

# Ceph v15.2 Document



Ceph是一个统一的分布式存储系统，设计初衷是提供较好的性能、可靠性和可扩展性。



下载手机APP  
畅享精彩阅读

# 目 录

致谢

Welcome

Intro to Ceph

Installing Ceph

Cephadm

Stability

Deploying a new Ceph cluster

Converting an existing cluster to cephadm

Upgrading Ceph

Cephadm operations

Cephadm monitoring

Client Setup

DriveGroups

Troubleshooting

Cephadm Concepts

Cephadm Feature Planning

Host Maintenance

Ceph Storage Cluster

Ceph File System

Ceph Block Device

Snapshots

Exclusive Locking

Mirroring

Live-Migration

Persistent Cache

Config Settings (librbd)

RBD Replay

Kernel Modules

QEMU

libvirt

Kubernetes

OpenStack

CloudStack

LIO iSCSI Gateway

Requirements

Configuring the iSCSI Target

[Using Ansible](#)

[Using the Command Line Interface](#)

[Configuring the iSCSI Initiators](#)

[Monitoring the iSCSI Gateways](#)

[rbd](#)

[rbd-fuse](#)

[rbd-nbd](#)

[rbd-ggate](#)

[rbd-map](#)

[ceph-rbdnamer](#)

[rbd-replay-prep](#)

[rbd-replay](#)

[rbd-replay-many](#)

[Ceph Object Gateway](#)

[HTTP Frontends](#)

[Pool Placement and Storage Classes](#)

[Multisite Configuration](#)

[Multisite Sync Policy Configuration](#)

[Configuring Pools](#)

[Config Reference](#)

[Admin Guide](#)

[S3 API](#)

[Common](#)

[Authentication](#)

[Service Ops](#)

[Bucket Ops](#)

[Object Ops](#)

[C++](#)

[C#](#)

[Java](#)

[Perl](#)

[PHP](#)

[Python](#)

[Ruby AWS::SDK Examples \(aws-sdk gem ~>2\)](#)

[Data caching and CDN](#)

[Swift API](#)

[Authentication](#)

[Service Ops](#)

- Container Ops
- Object Ops
- Temp URL Ops

Tutorial

- Java
- Python
- Ruby

Admin Ops API

Export over NFS

OpenStack Keystone Integration

OpenStack Barbican Integration

HashiCorp Vault Integration

Open Policy Agent Integration

Multi-tenancy

Compression

LDAP Authentication

Server-Side Encryption

Bucket Policy

Dynamic bucket index resharding

Multi factor authentication

Sync Modules

- ElasticSearch Sync Module
- Cloud Sync Module
- PubSub Module
- Archive Sync Module

Bucket Notifications

- S3 Bucket Notification Compatibility

Data Layout in RADOS

STS

STS Lite

Keycloak

Role

Orphan List and Associated Tooling

OpenID Connect Provider

Troubleshooting

Manpage radosgw

Manpage radosgw-admin

QAT Acceleration for Encryption and Compression

S3-select  
Lua Scripting  
Ceph Manager Daemon

Installation and Configuration

Writing modules  
Writing orchestrator plugins  
Ceph RESTful API  
Alerts module  
DiskPrediction module  
Local pool module  
RESTful module  
Zabbix module  
Prometheus module  
Influx module  
Hello module  
Telegraf module  
Telemetry module  
Iostat module  
Crash module  
Insights module  
Orchestrator module  
Rook module  
MDS Autoscaler module

Ceph Dashboard  
API Documentation

Architecture  
Developer Guide

Introduction  
Essentials  
What is Merged and When  
Issue tracker  
Basic workflow  
Tests: Unit Tests  
Tests: Integration Tests  
Running Tests Locally  
Running Integration Tests using Teuthology  
Running Tests in the Cloud  
Ceph Dashboard Developer Documentation (formerly HACKING.rst)

[UI Design Goals](#)

[Ceph Internals](#)

[Governance](#)

[Ceph Foundation](#)

[ceph-volume](#)

[Ceph Releases \(general\)](#)

[Ceph Releases \(index\)](#)

[v15.2.7 Octopus](#)

[v14.2.15 Nautilus](#)

[Archived releases index](#)

[v13.2.10 Mimic](#)

[v12.2.13 Luminous](#)

[v11.2.1 Kraken](#)

[v10.2.11 Jewel](#)

[v9.2.1 Infernalis](#)

[v0.94.10 Hammer](#)

[v0.87.2 Giant](#)

[v0.80.11 Firefly](#)

[v0.72.3 Emperor \(pending release\)](#)

[v0.67.12 “Dumpling” \(draft\)](#)

[v0.61.9 “Cuttlefish”](#)

[v0.56.7 “bobtail”](#)

[v0.48.3 “argonaut”](#)

[Glossary](#)

# 致谢

当前文档《Ceph v15.2 Document》由进击的皇虫使用书栈网(BookStack.CN)进行构建，生成于2020-12-05。

书栈网仅提供文档编写、整理、归类等功能，以及对文档内容的生成和导出工具。

文档内容由网友们编写和整理，书栈网难以确认文档内容知识点是否错漏。如果您在阅读文档获取知识的时候，发现文档内容有不恰当的地方，请向我们反馈，让我们共同携手，将知识准确、高效且有效地传递给每一个人。

同时，如果您在日常工作、生活和学习中遇到有价值有营养的知识文档，欢迎分享到书栈网，为知识的传承献上您的一份力量！

如果当前文档生成时间太久，请到书栈网获取最新的文档，以跟上知识更新换代的步伐。

内容来源：[Ceph](https://docs.ceph.com/en/latest/) <https://docs.ceph.com/en/latest/>

文档地址：<http://www.bookstack.cn/books/ceph-15.2-en>

书栈官网：<https://www.bookstack.cn>

书栈开源：<https://github.com/TruthHun>

分享，让知识传承更久远！感谢知识的创造者，感谢知识的分享者，也感谢每一位阅读到此处的读者，因为我们都将成为知识的传承者。

# Welcome to Ceph

Ceph uniquely delivers **object, block, and file storage in one unified system**.

Ceph Object Store	Ceph Block Device	Ceph File System
<ul style="list-style-type: none"><li>• RESTful Interface</li><li>• S3- and Swift-compliant APIs</li><li>• S3-style subdomains</li><li>• Unified S3/Swift namespace</li><li>• User management</li><li>• Usage tracking</li><li>• Striped objects</li><li>• Cloud solution integration</li><li>• Multi-site deployment</li><li>• Multi-site replication</li></ul>	<ul style="list-style-type: none"><li>• Thin-provisioned</li><li>• Images up to 16 exabytes</li><li>• Configurable striping</li><li>• In-memory caching</li><li>• Snapshots</li><li>• Copy-on-write cloning</li><li>• Kernel driver support</li><li>• KVM/libvirt support</li><li>• Back-end for cloud solutions</li><li>• Incremental backup</li><li>• Disaster recovery (multisite asynchronous replication)</li></ul>	<ul style="list-style-type: none"><li>• POSIX-compliant semantics</li><li>• Separates metadata from data</li><li>• Dynamic rebalancing</li><li>• Subdirectory snapshots</li><li>• Configurable striping</li><li>• Kernel driver support</li><li>• FUSE support</li><li>• NFS/CIFS deployable</li><li>• Use with Hadoop (replace HDFS)</li></ul>
See <a href="#">Ceph Object Store</a> for additional details.	See <a href="#">Ceph Block Device</a> for additional details.	See <a href="#">Ceph File System</a> for additional details.

Ceph is highly reliable, easy to manage, and free. The power of Ceph can transform your company's IT infrastructure and your ability to manage vast amounts of data. To try Ceph, see our [Getting Started](#) guides. To learn more about Ceph, see our [Architecture](#) section.

# Intro to Ceph

Whether you want to provide [Ceph Object Storage](#) and/or [Ceph Block Device](#) services to [Cloud Platforms](#), deploy a [Ceph File System](#) or use Ceph for another purpose, all [Ceph Storage Cluster](#) deployments begin with setting up each [Ceph Node](#), your network, and the Ceph Storage Cluster. A Ceph Storage Cluster requires at least one Ceph Monitor, Ceph Manager, and Ceph OSD (Object Storage Daemon). The Ceph Metadata Server is also required when running Ceph File System clients.



- **Monitors:** A [Ceph Monitor](#) (`ceph-mon`) maintains maps of the cluster state, including the monitor map, manager map, the OSD map, the MDS map, and the CRUSH map. These maps are critical cluster state required for Ceph daemons to coordinate with each other. Monitors are also responsible for managing authentication between daemons and clients. At least three monitors are normally required for redundancy and high availability.
- **Managers:** A [Ceph Manager](#) daemon (`ceph-mgr`) is responsible for keeping track of runtime metrics and the current state of the Ceph cluster, including storage utilization, current performance metrics, and system load. The Ceph Manager daemons also host python-based modules to manage and expose Ceph cluster information, including a web-based [Ceph Dashboard](#) and [REST API](#). At least two managers are normally required for high availability.
- **Ceph OSDs:** A [Ceph OSD](#) (object storage daemon, `ceph-osd`) stores data, handles data replication, recovery, rebalancing, and provides some monitoring information to Ceph Monitors and Managers by checking other Ceph OSD Daemons for a heartbeat. At least 3 Ceph OSDs are normally required for redundancy and high availability.
- **MDSs:** A [Ceph Metadata Server](#) (MDS, `ceph-mds`) stores metadata on behalf of the [Ceph File System](#) (i.e., Ceph Block Devices and Ceph Object Storage do not use MDS). Ceph Metadata Servers allow POSIX file system users to execute basic commands (like `ls`, `find`, etc.) without placing an enormous burden on the Ceph Storage Cluster.

Ceph stores data as objects within logical storage pools. Using the [CRUSH](#) algorithm, Ceph calculates which placement group should contain the object, and further calculates which Ceph OSD Daemon should store the placement group. The CRUSH algorithm enables the Ceph Storage Cluster to scale, rebalance, and recover dynamically.

## Recommendations

To begin using Ceph in production, you should review our hardware recommendations and operating system recommendations.

- [Hardware Recommendations](#)
  - [CPU](#)
  - [RAM](#)
  - [Memory](#)
  - [Data Storage](#)
  - [Networks](#)
  - [Failure Domains](#)
  - [Minimum Hardware Recommendations](#)
- [OS Recommendations](#)
  - [Ceph Dependencies](#)
  - [Platforms](#)

## Get Involved

You can avail yourself of help or contribute documentation, source code or bugs by getting involved in the Ceph community.

- [Get Involved in the Ceph Community!](#)
- [Documenting Ceph](#)
  - [Making Contributions](#)
  - [Documentation Style Guide](#)

# Installing Ceph

There are several different ways to install Ceph. Choose the method that best suits your needs.

## Recommended methods

[cephadm](#) installs and manages a Ceph cluster using containers and systemd, with tight integration with the CLI and dashboard GUI.

- `cephadm` only supports Octopus and newer releases.
- `cephadm` is fully integrated with the new orchestration API and fully supports the new CLI and dashboard features to manage cluster deployment.
- `cephadm` requires container support (podman or docker) and Python 3.

[Rook](#) deploys and manages Ceph clusters running in Kubernetes, while also enabling management of storage resources and provisioning via Kubernetes APIs. We recommend Rook as the way to run Ceph in Kubernetes or to connect an existing Ceph storage cluster to Kubernetes.

- Rook only supports Nautilus and newer releases of Ceph.
- Rook is the preferred method for running Ceph on Kubernetes, or for connecting a Kubernetes cluster to an existing (external) Ceph cluster.
- Rook supports the new orchestrator API. New management features in the CLI and dashboard are fully supported.

## Other methods

[ceph-ansible](#) deploys and manages Ceph clusters using Ansible.

- `ceph-ansible` is widely deployed.
- `ceph-ansible` is not integrated with the new orchestrator APIs, introduced in Nautilus and Octopus, which means that newer management features and dashboard integration are not available.

[ceph-deploy](#) is a tool for quickly deploying clusters.

### Important

`ceph-deploy` is no longer actively maintained. It is not tested on versions of Ceph newer than Nautilus. It does not support RHEL8, CentOS 8, or newer operating systems.

[DeepSea](#) installs Ceph using Salt.

[jaas.ai/ceph-mon](#) installs Ceph using Juju.

[github.com/openstack/puppet-ceph](#) installs Ceph via Puppet.

Ceph can also be [installed manually](#).

# Cephadm

---

Cephadm deploys and manages a Ceph cluster by connection to hosts from the manager daemon via SSH to add, remove, or update Ceph daemon containers. It does not rely on external configuration or orchestration tools like Ansible, Rook, or Salt.

Cephadm manages the full lifecycle of a Ceph cluster. It starts by bootstrapping a tiny Ceph cluster on a single node (one monitor and one manager) and then uses the orchestration interface (“day 2” commands) to expand the cluster to include all hosts and to provision all Ceph daemons and services. This can be performed via the Ceph command-line interface (CLI) or dashboard (GUI).

Cephadm is new in the Octopus v15.2.0 release and does not support older versions of Ceph.

## Note

Cephadm is new. Please read about [Stability](#) before using cephadm to deploy a production system.

- [Stability](#)
- [Deploying a new Ceph cluster](#)
- [Converting an existing cluster to cephadm](#)
- [Upgrading Ceph](#)
- [Cephadm operations](#)
- [Cephadm monitoring](#)
- [Cephadm CLI](#)
- [Client Setup](#)
- [DriveGroups](#)
- [Troubleshooting](#)
- [Cephadm Concepts](#)
- [Cephadm Feature Planning](#)

# Stability

Cephadm is a new feature in the Octopus release and has seen limited use in production and at scale. We would like users to try cephadm, especially for new clusters, but please be aware that some functionality is still rough around the edges. We expect fairly frequent updates and improvements over the first several bug fix releases of Octopus.

Cephadm management of the following components are currently well-supported:

- Monitors
- Managers
- OSDs
- CephFS file systems
- rbd-mirror

The following components are working with cephadm, but the documentation is not as complete as we would like, and there may be some changes in the near future:

- RGW
- dmcrypt OSDs

Cephadm support for the following features is still under development:

- NFS
- iSCSI

If you run into problems, you can always pause cephadm with:

```
1. ceph orch pause
```

Or turn cephadm off completely with:

```
1. ceph orch set backend ''
2. ceph mgr module disable cephadm
```

# Deploying a new Ceph cluster

Cephadm creates a new Ceph cluster by “bootstrapping” on a single host, expanding the cluster to encompass any additional hosts, and then deploying the needed services.

## Requirements

- Systemd
- Podman or Docker for running containers
- Time synchronization (such as chrony or NTP)
- LVM2 for provisioning storage devices

Any modern Linux distribution should be sufficient. Dependencies are installed automatically by the bootstrap process below.

## Install cephadm

The `cephadm` command can

1. bootstrap a new cluster
2. launch a containerized shell with a working Ceph CLI, and
3. aid in debugging containerized Ceph daemons.

There are a few ways to install `cephadm`:

- Use `curl` to fetch the most recent version of the standalone script.

```
1. curl --silent --remote-name --location https://github.com/ceph/ceph/raw/octopus/src/cephadm/cephadm
```

Make the `cephadm` script executable:

```
1. chmod +x cephadm
```

This script can be run directly from the current directory:

```
1. ./cephadm <arguments...>
```

- Although the standalone script is sufficient to get a cluster started, it is convenient to have the `cephadm` command installed on the host. To install the packages that provide the `cephadm` command for the Octopus release, run the

following commands:

```
1. ./cephadm add-repo --release octopus  
2. ./cephadm install
```

Confirm that `cephadm` is now in your PATH by running `which`:

```
1. which cephadm
```

A successful `which cephadm` command will return this:

```
1. /usr/sbin/cephadm
```

- Some commercial Linux distributions (e.g., RHEL, SLE) may already include up-to-date Ceph packages. In that case, you can install `cephadm` directly. For example:

```
1. dnf install -y cephadm
```

or

```
1. zypper install -y cephadm
```

## Bootstrap a new cluster

You need to know which *IP address* to use for the cluster's first monitor daemon. This is normally just the IP for the first host. If there are multiple networks and interfaces, be sure to choose one that will be accessible by any host accessing the Ceph cluster.

To bootstrap the cluster, first create an `/etc/ceph` directory:

```
1. mkdir -p /etc/ceph
```

Then run the `ceph bootstrap` command:

```
1. cephadm bootstrap --mon-ip *<mon-ip>*
```

This command will:

- Create a monitor and manager daemon for the new cluster on the local host.
- Generate a new SSH key for the Ceph cluster and adds it to the root user's `/root/.ssh/authorized_keys` file.
- Write a minimal configuration file needed to communicate with the new cluster to

- Write a copy of the `client.admin` administrative (privileged!) secret key to `/etc/ceph/ceph.client.admin.keyring`.
- Write a copy of the public key to `/etc/ceph/ceph.pub`.

The default bootstrap behavior will work for the vast majority of users. See below for a few options that may be useful for some users, or run `cephadm bootstrap -h` to see all available options:

- Bootstrap writes the files needed to access the new cluster to `/etc/ceph`, so that any Ceph packages installed on the host itself (e.g., to access the command line interface) can easily find them.

Daemon containers deployed with `cephadm`, however, do not need `/etc/ceph` at all. Use the `--output-dir *<directory>*` option to put them in a different directory (like `.`), avoiding any potential conflicts with existing Ceph configuration (`cephadm` or otherwise) on the same host.

- You can pass any initial Ceph configuration options to the new cluster by putting them in a standard ini-style configuration file and using the `--config *<config-file>*` option.
- You can choose the ssh user `cephadm` will use to connect to hosts by using the `--ssh-user *<user>*` option. The ssh key will be added to `/home/*<user>*/.ssh/authorized_keys`. This user will require passwordless sudo access.
- If you are using a container on an authenticated registry that requires login you may add the three arguments `--registry-url <url of registry>`, `--registry-username <username of account on registry>`, `--registry-password <password of account on registry>` OR `--registry-json <json file with login info>`. `Cephadm` will attempt to login to this registry so it may pull your container and then store the login info in its config database so other hosts added to the cluster may also make use of the authenticated registry.

## Enable Ceph CLI

---

`Cephadm` does not require any Ceph packages to be installed on the host. However, we recommend enabling easy access to the `ceph` command. There are several ways to do this:

- The `cephadm shell` command launches a bash shell in a container with all of the Ceph packages installed. By default, if configuration and keyring files are found in `/etc/ceph` on the host, they are passed into the container environment so that the shell is fully functional. Note that when executed on a MON host, `cephadm shell` will infer the `config` from the MON container instead of using the default configuration. If `--mount <path>` is given, then the host `<path>` (file or directory) will appear under `/mnt` inside the container:

```
1. cephadm shell
```

- To execute `ceph` commands, you can also run commands like this:

```
1. cephadm shell -- ceph -s
```

- You can install the `ceph-common` package, which contains all of the ceph commands, including `ceph`, `rbd`, `mount.ceph` (for mounting CephFS file systems), etc.:

```
1. cephadm add-repo --release octopus  
2. cephadm install ceph-common
```

Confirm that the `ceph` command is accessible with:

```
1. ceph -v
```

Confirm that the `ceph` command can connect to the cluster and also its status with:

```
1. ceph status
```

## Add hosts to the cluster

To add each new host to the cluster, perform two steps:

1. Install the cluster's public SSH key in the new host's root user's `authorized_keys` file:

```
1. ssh-copy-id -f -i /etc/ceph/ceph.pub root@*<new-host>*
```

For example:

```
1. ssh-copy-id -f -i /etc/ceph/ceph.pub root@host2  
2. ssh-copy-id -f -i /etc/ceph/ceph.pub root@host3
```

2. Tell Ceph that the new node is part of the cluster:

```
1. ceph orch host add *newhost*
```

For example:

```
1. ceph orch host add host2  
2. ceph orch host add host3
```

## Deploy additional monitors (optional)

A typical Ceph cluster has three or five monitor daemons spread across different hosts. We recommend deploying five monitors if there are five or more nodes in your cluster.

When Ceph knows what IP subnet the monitors should use it can automatically deploy and scale monitors as the cluster grows (or contracts). By default, Ceph assumes that other monitors should use the same subnet as the first monitor's IP.

If your Ceph monitors (or the entire cluster) live on a single subnet, then by default cephadm automatically adds up to 5 monitors as you add new hosts to the cluster. No further steps are necessary.

- If there is a specific IP subnet that should be used by monitors, you can configure that in `CIDR` format (e.g., `10.1.2.0/24`) with:

```
1. ceph config set mon public_network *<mon-cidr-network>*
```

For example:

```
1. ceph config set mon public_network 10.1.2.0/24
```

Cephadm deploys new monitor daemons only on hosts that have IPs configured in the configured subnet.

- If you want to adjust the default of 5 monitors, run this command:

```
1. ceph orch apply mon *<number-of-monitors>*
```

- To deploy monitors on a specific set of hosts, run this command:

```
1. ceph orch apply mon *<host1,host2,host3,...*>*
```

Be sure to include the first (bootstrap) host in this list.

- You can control which hosts the monitors run on by making use of host labels. To set the `mon` label to the appropriate hosts, run this command:

```
1. ceph orch host label add *<hostname>* mon
```

To view the current hosts and labels, run this command:

```
1. ceph orch host ls
```

For example:

```
1. ceph orch host label add host1 mon  
2. ceph orch host label add host2 mon  
3. ceph orch host label add host3 mon  
4. ceph orch host ls
```

	HOST	ADDR	LABELS	STATUS
2.	host1		mon	
3.	host2		mon	
4.	host3		mon	
5.	host4			
6.	host5			

Tell cephadm to deploy monitors based on the label by running this command:

```
1. ceph orch apply mon label:mon
```

- You can explicitly specify the IP address or CIDR network for each monitor and control where it is placed. To disable automated monitor deployment, run this command:

```
1. ceph orch apply mon --unmanaged
```

To deploy each additional monitor:

```
1. ceph orch daemon add mon *<host1:ip-or-network1> [<host1:ip-or-network-2>...]*
```

For example, to deploy a second monitor on `newhost1` using an IP address `10.1.2.123` and a third monitor on `newhost2` in network `10.1.2.0/24`, run the following commands:

```
1. ceph orch apply mon --unmanaged  
2. ceph orch daemon add mon newhost1:10.1.2.123  
3. ceph orch daemon add mon newhost2:10.1.2.0/24
```

## Note

The `apply` command can be confusing. For this reason, we recommend using YAML specifications.

Each `ceph orch apply mon` command supersedes the one before it. This means that you must use the proper comma-separated list-based syntax when you want to apply monitors to more than one host. If you do not use the proper syntax, you will clobber your work as you go.

For example:

```
1. ceph orch apply mon host1
2. ceph orch apply mon host2
3. ceph orch apply mon host3
```

This results in only one host having a monitor applied to it: host 3.

(The first command creates a monitor on host1. Then the second command clobbers the monitor on host1 and creates a monitor on host2. Then the third command clobbers the monitor on host2 and creates a monitor on host3. In this scenario, at this point, there is a monitor ONLY on host3.)

To make certain that a monitor is applied to each of these three hosts, run a command like this:

```
1. ceph orch apply mon "host1,host2,host3"
```

There is another way to apply monitors to multiple hosts: a `yaml` file can be used. Instead of using the “ceph orch apply mon” commands, run a command of this form:

```
1. ceph orch apply -i file.yaml
```

Here is a sample `file.yaml` file:

```
1. service_type: mon
2. placement:
3.   hosts:
4.     - host1
5.     - host2
6.     - host3
```

## Deploy OSDs

An inventory of storage devices on all cluster hosts can be displayed with:

```
1. ceph orch device ls
```

A storage device is considered *available* if all of the following conditions are met:

- The device must have no partitions.
- The device must not have any LVM state.
- The device must not be mounted.
- The device must not contain a file system.

- The device must not contain a Ceph BlueStore OSD.
- The device must be larger than 5 GB.

Ceph refuses to provision an OSD on a device that is not available.

There are a few ways to create new OSDs:

- Tell Ceph to consume any available and unused storage device:

```
1. ceph orch apply osd --all-available-devices
```

- Create an OSD from a specific device on a specific host:

```
1. ceph orch daemon add osd *<host>*<device-path>*
```

For example:

```
1. ceph orch daemon add osd host1:/dev/sdb
```

- Use [OSD Service Specification](#) to describe device(s) to consume based on their properties, such device type (SSD or HDD), device model names, size, or the hosts on which the devices exist:

```
1. ceph orch apply osd -i spec.yml
```

## Deploy MDSs

One or more MDS daemons is required to use the CephFS file system. These are created automatically if the newer `ceph fs volume` interface is used to create a new file system. For more information, see [FS volumes and subvolumes](#).

To deploy metadata servers:

```
1. ceph orch apply mds *<fs-name>* --placement="*<num-daemons>* [<host1>* ...]"
```

See [Placement Specification](#) for details of the placement specification.

## Deploy RGWs

Cephadm deploys radosgw as a collection of daemons that manage a particular *realm* and *zone*. (For more information about realms and zones, see [Multi-Site](#).)

Note that with cephadm, radosgw daemons are configured via the monitor configuration database instead of via a `ceph.conf` or the command line. If that configuration isn't

already in place (usually in the `client.rgw.<realmname>.<zonenumber>` section), then the radosgw daemons will start up with default settings (e.g., binding to port 80).

To deploy a set of radosgw daemons for a particular realm and zone, run the following command:

```
1. ceph orch apply rgw *<realm-name>* *<zone-name>* --placement="*<num-daemons>* [*<host1>* ...]"
```

For example, to deploy 2 rgw daemons serving the *myorg* realm and the *us-east-1* zone on *myhost1* and *myhost2*:

```
1. ceph orch apply rgw myorg us-east-1 --placement="2 myhost1 myhost2"
```

Cephadm will wait for a healthy cluster and automatically create the supplied realm and zone if they do not exist before deploying the rgw daemon(s)

Alternatively, the realm, zonegroup, and zone can be manually created using `radosgw-admin` commands:

```
1. radosgw-admin realm create --rgw-realm=<realm-name> --default
```

```
1. radosgw-admin zonegroup create --rgw-zonegroup=<zonegroup-name> --master --default
```

```
1. radosgw-admin zone create --rgw-zonegroup=<zonegroup-name> --rgw-zone=<zone-name> --master --default
```

```
1. radosgw-admin period update --rgw-realm=<realm-name> --commit
```

See [Placement Specification](#) for details of the placement specification.

## Deploying NFS ganesha

Cephadm deploys NFS Ganesha using a pre-defined RADOS pool and optional *namespace*

To deploy a NFS Ganesha gateway, run the following command:

```
1. ceph orch apply nfs *<svc_id>* *<pool>* *<namespace>* --placement="*<num-daemons>* [*<host1>* ...]"
```

For example, to deploy NFS with a service id of *foo*, that will use the RADOS pool *nfs-ganesha* and namespace *nfs-ns*:

```
1. ceph orch apply nfs foo nfs-ganesha nfs-ns
```

### Note

Create the *nfs-ganesha* pool first if it doesn't exist.

See [Placement Specification](#) for details of the placement specification.

## Deploying custom containers

---

It is also possible to choose different containers than the default containers to deploy Ceph. See [Ceph Container Images](#) for information about your options in this regard.

# Converting an existing cluster to cephadm

Cephadm allows you to convert an existing Ceph cluster that has been deployed with ceph-deploy, ceph-ansible, DeepSea, or similar tools.

## Limitations

- Cephadm only works with BlueStore OSDs. If there are FileStore OSDs in your cluster you cannot manage them.

## Preparation

1. Get the `cephadm` command line tool on each host in the existing cluster. See [Install cephadm](#).

2. Prepare each host for use by `cephadm` :

```
1. # cephadm prepare-host
```

3. Determine which Ceph version you will use. You can use any Octopus (15.2.z) release or later. For example, `docker.io/ceph/ceph:v15.2.0`. The default will be the latest stable release, but if you are upgrading from an earlier release at the same time be sure to refer to the upgrade notes for any special steps to take while upgrading.

The image is passed to cephadm with:

```
1. # cephadm --image $IMAGE <rest of command goes here>
```

4. Cephadm can provide a list of all Ceph daemons on the current host:

```
1. # cephadm ls
```

Before starting, you should see that all existing daemons have a style of `legacy` in the resulting output. As the adoption process progresses, adopted daemons will appear as style `cephadm:v1`.

## Adoption process

1. Ensure the ceph configuration is migrated to use the cluster config database. If the `/etc/ceph/ceph.conf` is identical on each host, then on one host:

```
1. # ceph config assimilate-conf -i /etc/ceph/ceph.conf
```

If there are config variations on each host, you may need to repeat this command on each host. You can view the cluster's configuration to confirm that it is complete with:

```
1. # ceph config dump
```

## 2. Adopt each monitor:

```
1. # cephadm adopt --style legacy --name mon.<hostname>
```

Each legacy monitor should stop, quickly restart as a cephadm container, and rejoin the quorum.

## 3. Adopt each manager:

```
1. # cephadm adopt --style legacy --name mgr.<hostname>
```

## 4. Enable cephadm:

```
1. # ceph mgr module enable cephadm  
2. # ceph orch set backend cephadm
```

## 5. Generate an SSH key:

```
1. # ceph cephadm generate-key  
2. # ceph cephadm get-pub-key > ~/ceph.pub
```

## 6. Install the cluster SSH key on each host in the cluster:

```
1. # ssh-copy-id -f -i ~/ceph.pub root@<host>
```

### Note

It is also possible to import an existing ssh key. See [ssh errors](#) in the troubleshooting document for instructions describing how to import existing ssh keys.

## 7. Tell cephadm which hosts to manage:

```
1. # ceph orch host add <hostname> [ip-address]
```

This will perform a `cephadm check-host` on each host before adding it to ensure it is working. The IP address argument is only required if DNS does not allow you to connect to each host by its short name.

## 8. Verify that the adopted monitor and manager daemons are visible:

```
1. # ceph orch ps
```

## 9. Adopt all OSDs in the cluster:

```
1. # cephadm adopt --style legacy --name <name>
```

For example:

```
1. # cephadm adopt --style legacy --name osd.1
2. # cephadm adopt --style legacy --name osd.2
```

## 10. Redeploy MDS daemons by telling cephadm how many daemons to run for each file system. You can list file systems by name with `ceph fs ls`. Run the following command on the master nodes:

```
1. # ceph orch apply mds <fs-name> [--placement=<placement>]
```

For example, in a cluster with a single file system called foo:

```
1. # ceph fs ls
2. name: foo, metadata pool: foo_metadata, data pools: [foo_data ]
3. # ceph orch apply mds foo 2
```

Wait for the new MDS daemons to start with:

```
1. # ceph orch ps --daemon-type mds
```

Finally, stop and remove the legacy MDS daemons:

```
1. # systemctl stop ceph-mds.target
2. # rm -rf /var/lib/ceph/mds/ceph-*
```

## 11. Redeploy RGW daemons. Cephadm manages RGW daemons by zone. For each zone, deploy new RGW daemons with cephadm:

```
# ceph orch apply rgw <realm> <zone> [--subcluster=<subcluster>] [--port=<port>] [--ssl] [--placement=<placement>]
```

where *<placement>* can be a simple daemon count, or a list of specific hosts (see [Placement Specification](#)).

Once the daemons have started and you have confirmed they are functioning, stop and remove the old legacy daemons:

```
1. # systemctl stop ceph-rgw.target
2. # rm -rf /var/lib/ceph/radosgw/ceph-*
```

For adopting single-site systems without a realm, see also [Migrating a Single Site System to Multi-Site](#).

12. Check the `ceph health detail` output for cephadm warnings about stray cluster daemons or hosts that are not yet managed.

# Upgrading Ceph

Cephadm is capable of safely upgrading Ceph from one bugfix release to another. For example, you can upgrade from v15.2.0 (the first Octopus release) to the next point release v15.2.1.

The automated upgrade process follows Ceph best practices. For example:

- The upgrade order starts with managers, monitors, then other daemons.
- Each daemon is restarted only after Ceph indicates that the cluster will remain available.

Keep in mind that the Ceph cluster health status is likely to switch to HEALTH\_WARNING during the upgrade.

## Starting the upgrade

Before you start, you should verify that all hosts are currently online and your cluster is healthy.

```
1. # ceph -s
```

To upgrade (or downgrade) to a specific release:

```
1. # ceph orch upgrade start --ceph-version <version>
```

For example, to upgrade to v15.2.1:

```
1. # ceph orch upgrade start --ceph-version 15.2.1
```

## Monitoring the upgrade

Determine whether an upgrade is in process and what version the cluster is upgrading to with:

```
1. # ceph orch upgrade status
```

While the upgrade is underway, you will see a progress bar in the ceph status output. For example:

```
1. # ceph -s
2. [...]
3.   progress:
```

```
4.      Upgrade to docker.io/ceph/ceph:v15.2.1 (00h 20m 12s)
5.      [=====.....] (time remaining: 01h 43m 31s)
```

You can also watch the cephadm log with:

```
1. # ceph -W cephadm
```

## Canceling an upgrade

You can stop the upgrade process at any time with:

```
1. # ceph orch upgrade stop
```

## Potential problems

There are a few health alerts that can arise during the upgrade process.

### UPGRADE\_NO\_STANDBY\_MGR

Ceph requires an active and standby manager daemon in order to proceed, but there is currently no standby.

You can ensure that Cephadm is configured to run 2 (or more) managers with:

```
1. # ceph orch apply mgr 2 # or more
```

You can check the status of existing mgr daemons with:

```
1. # ceph orch ps --daemon-type mgr
```

If an existing mgr daemon has stopped, you can try restarting it with:

```
1. # ceph orch daemon restart <name>
```

### UPGRADE\_FAILED\_PULL

Ceph was unable to pull the container image for the target version. This can happen if you specify an version or container image that does not exist (e.g., 1.2.3), or if the container registry is not reachable from one or more hosts in the cluster.

You can cancel the existing upgrade and specify a different target version with:

```
1. # ceph orch upgrade stop
2. # ceph orch upgrade start --ceph-version <version>
```

## Using customized container images

For most users, specifying the Ceph version is sufficient. Cephadm will locate the specific Ceph container image to use by combining the `container_image_base` configuration option (default: `docker.io/ceph/ceph`) with a tag of `vX.Y.Z`.

You can also upgrade to an arbitrary container image. For example, to upgrade to a development build:

```
1. # ceph orch upgrade start --image quay.io/ceph-ci/ceph:recent-git-branch-name
```

For more information about available container images, see [Ceph Container Images](#).

# Cephadm Operations

## Watching cephadm log messages

Cephadm logs to the `cephadm` cluster log channel, meaning you can monitor progress in realtime with:

```
1. # ceph -W cephadm
```

By default it will show info-level events and above. To see debug-level messages too:

```
1. # ceph config set mgr mgr/cephadm/log_to_cluster_level debug
2. # ceph -W cephadm --watch-debug
```

Be careful: the debug messages are very verbose!

You can see recent events with:

```
1. # ceph log last cephadm
```

These events are also logged to the `ceph.cephadm.log` file on monitor hosts and to the monitor daemons' stderr.

## Ceph daemon logs

### Logging to stdout

Traditionally, Ceph daemons have logged to `/var/log/ceph`. By default, cephadm daemons log to stderr and the logs are captured by the container runtime environment. For most systems, by default, these logs are sent to journald and accessible via `journalctl`.

For example, to view the logs for the daemon `mon.foo` for a cluster with ID `5c5a50ae-272a-455d-99e9-32c6a013e694`, the command would be something like:

```
1. journalctl -u ceph-5c5a50ae-272a-455d-99e9-32c6a013e694@mon.foo
```

This works well for normal operations when logging levels are low.

To disable logging to stderr:

```
1. ceph config set global log_to_stderr false
2. ceph config set global mon_cluster_log_to_stderr false
```

## Logging to files

You can also configure Ceph daemons to log to files instead of stderr, just like they have in the past. When logging to files, Ceph logs appear in `/var/log/ceph/<cluster-fsid>`.

To enable logging to files:

1. `ceph config set global log_to_file true`
2. `ceph config set global mon_cluster_log_to_file true`

We recommend disabling logging to stderr (see above) or else everything will be logged twice:

1. `ceph config set global log_to_stderr false`
2. `ceph config set global mon_cluster_log_to_stderr false`

By default, cephadm sets up log rotation on each host to rotate these files. You can configure the logging retention schedule by modifying `/etc/logrotate.d/ceph.<cluster-fsid>`.

## Data location

Cephadm daemon data and logs in slightly different locations than older versions of ceph:

- `/var/log/ceph/<cluster-fsid>` contains all cluster logs. Note that by default cephadm logs via stderr and the container runtime, so these logs are normally not present.
- `/var/lib/ceph/<cluster-fsid>` contains all cluster daemon data (besides logs).
- `/var/lib/ceph/<cluster-fsid>/<daemon-name>` contains all data for an individual daemon.
- `/var/lib/ceph/<cluster-fsid>/crash` contains crash reports for the cluster.
- `/var/lib/ceph/<cluster-fsid>/removed` contains old daemon data directories for stateful daemons (e.g., monitor, prometheus) that have been removed by cephadm.

## Disk usage

Because a few Ceph daemons may store a significant amount of data in `/var/lib/ceph` (notably, the monitors and prometheus), we recommend moving this directory to its own disk, partition, or logical volume so that it does not fill up the root file system.

## SSH Configuration

Cephadm uses SSH to connect to remote hosts. SSH uses a key to authenticate with those

hosts in a secure way.

## Default behavior

Cephadm stores an SSH key in the monitor that is used to connect to remote hosts. When the cluster is bootstrapped, this SSH key is generated automatically and no additional configuration is necessary.

A new SSH key can be generated with:

```
1. ceph cephadm generate-key
```

The public portion of the SSH key can be retrieved with:

```
1. ceph cephadm get-pub-key
```

The currently stored SSH key can be deleted with:

```
1. ceph cephadm clear-key
```

You can make use of an existing key by directly importing it with:

```
1. ceph config-key set mgr/cephadm/ssh_identity_key -i <key>
2. ceph config-key set mgr/cephadm/ssh_identity_pub -i <pub>
```

You will then need to restart the mgr daemon to reload the configuration with:

```
1. ceph mgr fail
```

## Configuring a different SSH user

Cephadm must be able to log into all the Ceph cluster nodes as an user that has enough privileges to download container images, start containers and execute commands without prompting for a password. If you do not want to use the “root” user (default option in cephadm), you must provide cephadm the name of the user that is going to be used to perform all the cephadm operations. Use the command:

```
1. ceph cephadm set-user <user>
```

Prior to running this the cluster ssh key needs to be added to this users authorized\_keys file and non-root users must have passwordless sudo access.

## Customizing the SSH configuration

Cephadm generates an appropriate `ssh_config` file that is used for connecting to remote

hosts. This configuration looks something like this:

```
1. Host *
2. User root
3. StrictHostKeyChecking no
4. UserKnownHostsFile /dev/null
```

There are two ways to customize this configuration for your environment:

1. Import a customized configuration file that will be stored by the monitor with:

```
1. ceph cephadm set-ssh-config -i <ssh_config_file>
```

To remove a customized SSH config and revert back to the default behavior:

```
1. ceph cephadm clear-ssh-config
```

2. You can configure a file location for the SSH configuration file with:

```
1. ceph config set mgr mgr/cephadm/ssh_config_file <path>
```

We do *not recommend* this approach. The path name must be visible to *any* mgr daemon, and cephadm runs all daemons as containers. That means that the file either need to be placed inside a customized container image for your deployment, or manually distributed to the mgr data directory (`/var/lib/ceph/<cluster-fsid>/mgr.<id>` on the host, visible at `/var/lib/ceph/mgr/ceph-<id>` from inside the container).

## Health checks

### CEPHADM\_PAUSED

Cephadm background work has been paused with `ceph orch pause`. Cephadm continues to perform passive monitoring activities (like checking host and daemon status), but it will not make any changes (like deploying or removing daemons).

Resume cephadm work with:

```
1. ceph orch resume
```

### CEPHADM\_STRAY\_HOST

One or more hosts have running Ceph daemons but are not registered as hosts managed by *cephadm*. This means that those services cannot currently be managed by cephadm (e.g., restarted, upgraded, included in `ceph orch ps`).

You can manage the host(s) with:

```
1. ceph orch host add *<hostname>*
```

Note that you may need to configure SSH access to the remote host before this will work.

Alternatively, you can manually connect to the host and ensure that services on that host are removed or migrated to a host that is managed by *cephadm*.

You can also disable this warning entirely with:

```
1. ceph config set mgr mgr/cephadm/warn_on_stray_hosts false
```

See [Fully qualified domain names vs bare host names](#) for more information about host names and domain names.

## CEPHADM\_STRAY\_DAEMON

One or more Ceph daemons are running but not are not managed by *cephadm*. This may be because they were deployed using a different tool, or because they were started manually. Those services cannot currently be managed by *cephadm* (e.g., restarted, upgraded, or included in *ceph orch ps*).

If the daemon is a stateful one (monitor or OSD), it should be adopted by *cephadm*; see [Converting an existing cluster to cephadm](#). For stateless daemons, it is usually easiest to provision a new daemon with the `ceph orch apply` command and then stop the unmanaged daemon.

This warning can be disabled entirely with:

```
1. ceph config set mgr mgr/cephadm/warn_on_stray_daemons false
```

## CEPHADM\_HOST\_CHECK\_FAILED

One or more hosts have failed the basic *cephadm* host check, which verifies that (1) the host is reachable and *cephadm* can be executed there, and (2) that the host satisfies basic prerequisites, like a working container runtime (podman or docker) and working time synchronization. If this test fails, *cephadm* will no be able to manage services on that host.

You can manually run this check with:

```
1. ceph cephadm check-host *<hostname>*
```

You can remove a broken host from management with:

```
1. ceph orch host rm *<hostname>*
```

You can disable this health warning with:

```
1. ceph config set mgr mgr/cephadm/warn_on_failed_host_check false
```

## /etc/ceph/ceph.conf

Cephadm distributes a minimized `ceph.conf` that only contains a minimal set of information to connect to the Ceph cluster.

To update the configuration settings, instead of manually editing the `ceph.conf` file, use the config database instead:

```
1. ceph config set ...
```

See [Monitor configuration database](#) for details.

By default, `cephadm` does not deploy that minimized `ceph.conf` across the cluster. To enable the management of `/etc/ceph/ceph.conf` files on all hosts, please enable this by running:

```
1. ceph config set mgr mgr/cephadm/manage_etc_ceph_ceph_conf true
```

To set up an initial configuration before bootstrapping the cluster, create an initial `ceph.conf` file. For example:

```
1. cat <<EOF > /etc/ceph/ceph.conf
2. [global]
3. osd crush chooseleaf type = 0
4. EOF
```

Then, run bootstrap referencing this file:

```
1. cephadm bootstrap -c /root/ceph.conf ...
```

# Monitoring Stack with Cephadm

Ceph Dashboard uses [Prometheus](#), [Grafana](#), and related tools to store and visualize detailed metrics on cluster utilization and performance. Ceph users have three options:

1. Have cephadm deploy and configure these services. This is the default when bootstrapping a new cluster unless the `--skip-monitoring-stack` option is used.
2. Deploy and configure these services manually. This is recommended for users with existing prometheus services in their environment (and in cases where Ceph is running in Kubernetes with Rook).
3. Skip the monitoring stack completely. Some Ceph dashboard graphs will not be available.

The monitoring stack consists of [Prometheus](#), Prometheus exporters ([Prometheus Module](#), [Node exporter](#)), [Prometheus Alert Manager](#) and [Grafana](#).

## Note

Prometheus' security model presumes that untrusted users have access to the Prometheus HTTP endpoint and logs. Untrusted users have access to all the (meta)data Prometheus collects that is contained in the database, plus a variety of operational and debugging information.

However, Prometheus' HTTP API is limited to read-only operations. Configurations can *not* be changed using the API and secrets are not exposed. Moreover, Prometheus has some built-in measures to mitigate the impact of denial of service attacks.

Please see Prometheus' Security model

<<https://prometheus.io/docs/operating/security/>>; for more detailed information.

## Deploying monitoring with cephadm

By default, bootstrap will deploy a basic monitoring stack. If you did not do this (by passing `--skip-monitoring-stack` , or if you converted an existing cluster to cephadm management, you can set up monitoring by following the steps below.

1. Enable the prometheus module in the ceph-mgr daemon. This exposes the internal Ceph metrics so that prometheus can scrape them.

```
1. ceph mgr module enable prometheus
```

2. Deploy a node-exporter service on every node of the cluster. The node-exporter provides host-level metrics like CPU and memory utilization.

```
1. ceph orch apply node-exporter '*'
```

### 3. Deploy alertmanager

```
1. ceph orch apply alertmanager 1
```

### 4. Deploy prometheus. A single prometheus instance is sufficient, but for HA you may want to deploy two.

```
1. ceph orch apply prometheus 1 # or 2
```

### 5. Deploy grafana

```
1. ceph orch apply grafana 1
```

Cephadm handles the prometheus, grafana, and alertmanager configurations automatically.

It may take a minute or two for services to be deployed. Once completed, you should see something like this from `ceph orch ls`

1. \$ ceph orch ls	2. NAME	RUNNING	REFRESHED	IMAGE NAME	IMAGE ID	SPEC
	3. alertmanager	1/1	6s ago	docker.io/prom/alertmanager:latest	0881eb8f169f	present
	4. crash	2/2	6s ago	docker.io/ceph/daemon-base:latest-master-devel	mix	present
	5. grafana	1/1	0s ago	docker.io/pcuzner/ceph-grafana-el8:latest	f77afcfc0bcf6	absent
	6. node-exporter	2/2	6s ago	docker.io/prom/node-exporter:latest	e5a616e4b9cf	present
	7. prometheus	1/1	6s ago	docker.io/prom/prometheus:latest	e935122ab143	present

## Configuring SSL/TLS for Grafana

`cephadm` will deploy Grafana using the certificate defined in the ceph key/value store. If a certificate is not specified, `cephadm` will generate a self-signed certificate during deployment of the Grafana service.

A custom certificate can be configured using the following commands.

```
1. ceph config-key set mgr/cephadm/grafana_key -i $PWD/key.pem
2. ceph config-key set mgr/cephadm/grafana_crt -i $PWD/certificate.pem
```

The `cephadm` manager module needs to be restarted to be able to read updates to these keys.

```
1. ceph orch restart mgr
```

If you already deployed Grafana, you need to redeploy the service for the configuration to be updated.

```
1. ceph orch redeploy grafana
```

The `redeploy` command also takes care of setting the right URL for Ceph Dashboard.

## Using custom images

It is possible to install or upgrade monitoring components based on other images. To do so, the name of the image to be used needs to be stored in the configuration first. The following configuration options are available.

- `container_image_prometheus`
- `container_image_grafana`
- `container_image_alertmanager`
- `container_image_node_exporter`

Custom images can be set with the `ceph config` command

```
1. ceph config set mgr mgr/cephadm/<option_name> <value>
```

For example

```
1. ceph config set mgr mgr/cephadm/container_image_prometheus prom/prometheus:v1.4.1
```

### Note

By setting a custom image, the default value will be overridden (but not overwritten). The default value changes when updates become available. By setting a custom image, you will not be able to update the component you have set the custom image for automatically. You will need to manually update the configuration (image name and tag) to be able to install updates.

If you choose to go with the recommendations instead, you can reset the custom image you have set before. After that, the default value will be used again. Use `ceph config rm` to reset the configuration option

```
1. ceph config rm mgr mgr/cephadm/<option_name>
```

For example

```
1. ceph config rm mgr mgr/cephadm/container_image_prometheus
```

# Using custom configuration files

By overriding cephadm templates, it is possible to completely customize the configuration files for monitoring services.

Internally, cephadm already uses `Jinja2` templates to generate the configuration files for all monitoring components. To be able to customize the configuration of Prometheus, Grafana or the Alertmanager it is possible to store a `Jinja2` template for each service that will be used for configuration generation instead. This template will be evaluated every time a service of that kind is deployed or reconfigured. That way, the custom configuration is preserved and automatically applied on future deployments of these services.

## Note

The configuration of the custom template is also preserved when the default configuration of cephadm changes. If the updated configuration is to be used, the custom template needs to be migrated *manually*.

## Option names

The following templates for files that will be generated by cephadm can be overridden. These are the names to be used when storing with `ceph config-key set` :

- `alertmanager_alertmanager.yml`
- `grafana_ceph-dashboard.yml`
- `grafana_grafana.ini`
- `prometheus_prometheus.yml`

You can look up the file templates that are currently used by cephadm in

`src/pybind/mgr/cephadm/templates` :

- `services/alertmanager/alertmanager.yml.j2`
- `services/grafana/ceph-dashboard.yml.j2`
- `services/grafana/grafana.ini.j2`
- `services/prometheus/prometheus.yml.j2`

## Usage

The following command applies a single line value:

```
1. ceph config-key set mgr/cephadm/<option_name> <value>
```

To set contents of files as template use the `-i` argument:

```
1. ceph config-key set mgr/cephadm/<option_name> -i $PWD/<filename>
```

## Note

When using files as input to `config-key` an absolute path to the file must be used.

It is required to restart the `cephadm` `mgr` module after a configuration option has been set. Then the configuration file for the service needs to be recreated. This is done using `redeploy`. For more details see the following example.

## Example

```
1. # set the contents of ./prometheus.yml.j2 as template
2. ceph config-key set mgr/cephadm/services_prometheus_prometheus.yml \
3.   -i $PWD/prometheus.yml.j2
4.
5. # restart cephadm mgr module
6. ceph orch restart mgr
7.
8. # redeploy the prometheus service
9. ceph orch redeploy prometheus
```

## Disabling monitoring

If you have deployed monitoring and would like to remove it, you can do so with

```
1. ceph orch rm grafana
2. ceph orch rm prometheus --force    # this will delete metrics data collected so far
3. ceph orch rm node-exporter
4. ceph orch rm alertmanager
5. ceph mgr module disable prometheus
```

## Deploying monitoring manually

If you have an existing Prometheus monitoring infrastructure, or would like to manage it yourself, you need to configure it to integrate with your Ceph cluster.

- Enable the Prometheus module in the `ceph-mgr` daemon

```
1. ceph mgr module enable prometheus
```

By default, `ceph-mgr` presents Prometheus metrics on port 9283 on each host running a `ceph-mgr` daemon. Configure Prometheus to scrape these.

- To enable the dashboard's Prometheus-based alerting, see [Enabling Prometheus Alerting](#).

- To enable dashboard integration with Grafana, see [Enabling the Embedding of Grafana Dashboards](#).

## Enabling RBD-Image monitoring

---

Due to performance reasons, monitoring of RBD images is disabled by default. For more information please see [RBD IO statistics](#). If disabled, the overview and details dashboards will stay empty in Grafana and the metrics will not be visible in Prometheus.

# Basic Ceph Client Setup

Client machines need some basic configuration in order to interact with a cluster. This document describes how to configure a client machine for cluster interaction.

## Note

Most client machines only need the ceph-common package and its dependencies installed. That will supply the basic ceph and rados commands, as well as other commands like mount.ceph and rbd.

## Config File Setup

Client machines can generally get away with a smaller config file than a full-fledged cluster member. To generate a minimal config file, log into a host that is already configured as a client or running a cluster daemon, and then run

```
1. ceph config generate-minimal-conf
```

This will generate a minimal config file that will tell the client how to reach the Ceph Monitors. The contents of this file should typically be installed in /etc/ceph/ceph.conf.

## Keyring Setup

Most Ceph clusters are run with authentication enabled, and the client will need keys in order to communicate with cluster machines. To generate a keyring file with credentials for client.fs, log into an extant cluster member and run

```
1. ceph auth get-or-create client.fs
```

The resulting output should be put into a keyring file, typically /etc/ceph/ceph.keyring.

# OSD Service Specification

**Service Specification** of type `osd` are a way to describe a cluster layout using the properties of disks. It gives the user an abstract way tell ceph which disks should turn into an OSD with which configuration without knowing the specifics of device names and paths.

Instead of doing this

```
1. ceph orch daemon add osd *<host>*:*<path-to-device>*
```

for each device and each host, we can define a yaml|json file that allows us to describe the layout. Here's the most basic example.

Create a file called i.e. `osd_spec.yml`

```
1. service_type: osd
2. service_id: default_drive_group <- name of the drive_group (name can be custom)
3. placement:
4.   host_pattern: '*' <- which hosts to target, currently only supports globs
5.   data_devices: <- the type of devices you are applying specs to
6.   all: true <- a filter, check below for a full list
```

This would translate to:

Turn any available(ceph-volume decides what 'available' is) into an OSD on all hosts that match the glob pattern '\*'. (The glob pattern matches against the registered hosts from host ls) There will be a more detailed section on host\_pattern down below.

and pass it to osd create like so

```
1. ceph orch apply osd -i /path/to/osd_spec.yml
```

This will go out on all the matching hosts and deploy these OSDs.

Since we want to have more complex setups, there are more filters than just the 'all' filter.

Also, there is a `-dry-run` flag that can be passed to the `apply osd` command, which gives you a synopsis of the proposed layout.

Example

```
1. [monitor.1]# ceph orch apply osd -i /path/to/osd_spec.yml --dry-run
```

# Filters

## Note

Filters are applied using a AND gate by default. This essentially means that a drive needs to fulfill all filter criteria in order to get selected. If you wish to change this behavior you can adjust this behavior by setting

```
filter_logic: OR # valid arguments are AND, OR
```

in the OSD Specification.

You can assign disks to certain groups by their attributes using filters.

The attributes are based off of ceph-volume's disk query. You can retrieve the information with

```
1. ceph-volume inventory </path/to/disk>
```

## Vendor or Model:

You can target specific disks by their Vendor or by their Model

```
1. model: disk_model_name
```

or

```
1. vendor: disk_vendor_name
```

## Size:

You can also match by disk Size.

```
1. size: size_spec
```

## Size specs:

Size specification of format can be of form:

- LOW:HIGH
- :HIGH
- LOW:
- EXACT

Concrete examples:

Includes disks of an exact size

```
1. size: '10G'
```

Includes disks which size is within the range

```
1. size: '10G:40G'
```

Includes disks less than or equal to 10G in size

```
1. size: ':10G'
```

Includes disks equal to or greater than 40G in size

```
1. size: '40G:'
```

Sizes don't have to be exclusively in Gigabyte(G).

Supported units are Megabyte(M), Gigabyte(G) and Terrabyte(T). Also appending the (B) for byte is supported. MB, GB, TB

## Rotational:

This operates on the 'rotational' attribute of the disk.

```
1. rotational: 0 | 1
```

1 to match all disks that are rotational

0 to match all disks that are non-rotational (SSD, NVME etc)

## All:

This will take all disks that are 'available'

Note: This is exclusive for the data\_devices section.

```
1. all: true
```

## Limiter:

When you specified valid filters but want to limit the amount of matching disks you can use the 'limit' directive.

```
1. limit: 2
```

For example, if you used vendor to match all disks that are from VendorA but only want to use the first two you could use limit.

```
1. data_devices:
2. vendor: VendorA
3. limit: 2
```

Note: Be aware that limit is really just a last resort and shouldn't be used if it can be avoided.

## Additional Options

There are multiple optional settings you can use to change the way OSDs are deployed. You can add these options to the base level of a DriveGroup for it to take effect.

This example would deploy all OSDs with encryption enabled.

```
1. service_type: osd
2. service_id: example_osd_spec
3. placement:
4. host_pattern: '*'
5. data_devices:
6. all: true
7. encrypted: true
```

See a full list in the DriveGroupSpecs

```
ceph.deployment.drive_group.``DriveGroupSpec
```

## Examples

### The simple case

All nodes with the same setup

```
1. 20 HDDs
2. Vendor: VendorA
3. Model: HDD-123-foo
4. Size: 4TB
5.
6. 2 SSDs
7. Vendor: VendorB
8. Model: MC-55-44-ZX
9. Size: 512GB
```

This is a common setup and can be described quite easily:

```

1. service_type: osd
2. service_id: osd_spec_default
3. placement:
4. host_pattern: '*'
5. data_devices:
6. model: HDD-123-foo <- note that HDD-123 would also be valid
7. db_devices:
8. model: MC-55-44-XZ <- same here, MC-55-44 is valid

```

However, we can improve it by reducing the filters on core properties of the drives:

```

1. service_type: osd
2. service_id: osd_spec_default
3. placement:
4. host_pattern: '*'
5. data_devices:
6. rotational: 1
7. db_devices:
8. rotational: 0

```

Now, we enforce all rotating devices to be declared as 'data devices' and all non-rotating devices will be used as shared\_devices (wal, db)

If you know that drives with more than 2TB will always be the slower data devices, you can also filter by size:

```

1. service_type: osd
2. service_id: osd_spec_default
3. placement:
4. host_pattern: '*'
5. data_devices:
6. size: '2TB:'
7. db_devices:
8. size: ':2TB'

```

Note: All of the above DriveGroups are equally valid. Which of those you want to use depends on taste and on how much you expect your node layout to change.

## The advanced case

Here we have two distinct setups

```

1. 20 HDDs
2. Vendor: VendorA
3. Model: HDD-123-foo
4. Size: 4TB
5.

```

```

6. 12 SSDs
7. Vendor: VendorB
8. Model: MC-55-44-ZX
9. Size: 512GB
10.
11. 2 NVMEs
12. Vendor: VendorC
13. Model: NVME-QQQQ-987
14. Size: 256GB

```

- 20 HDDs should share 2 SSDs

- 10 SSDs should share 2 NVMEs

This can be described with two layouts.

```

1. service_type: osd
2. service_id: osd_spec_hdd
3. placement:
4. host_pattern: '*'
5. data_devices:
6. rotational: 0
7. db_devices:
8. model: MC-55-44-XZ
9. limit: 2 (db_slots is actually to be favoured here, but it's not implemented yet)
10.
11. service_type: osd
12. service_id: osd_spec_ssd
13. placement:
14. host_pattern: '*'
15. data_devices:
16. model: MC-55-44-XZ
17. db_devices:
18. vendor: VendorC

```

This would create the desired layout by using all HDDs as data\_devices with two SSD assigned as dedicated db/wal devices. The remaining SSDs(8) will be data\_devices that have the 'VendorC' NVMEs assigned as dedicated db/wal devices.

## The advanced case (with non-uniform nodes)

The examples above assumed that all nodes have the same drives. That's however not always the case.

Node1-5

```

1. 20 HDDs
2. Vendor: Intel
3. Model: SSD-123-foo
4. Size: 4TB

```

```

5. 2 SSDs
6. Vendor: VendorA
7. Model: MC-55-44-ZX
8. Size: 512GB

```

## Node6-10

```

1. 5 NVMEs
2. Vendor: Intel
3. Model: SSD-123-foo
4. Size: 4TB
5. 20 SSDs
6. Vendor: VendorA
7. Model: MC-55-44-ZX
8. Size: 512GB

```

You can use the 'host\_pattern' key in the layout to target certain nodes. Salt target notation helps to keep things easy.

```

1. service_type: osd
2. service_id: osd_spec_node_one_to_five
3. placement:
4. host_pattern: 'node[1-5]'
5. data_devices:
6. rotational: 1
7. db_devices:
8. rotational: 0
9.
10.
11. service_type: osd
12. service_id: osd_spec_six_to_ten
13. placement:
14. host_pattern: 'node[6-10]'
15. data_devices:
16. model: MC-55-44-XZ
17. db_devices:
18. model: SSD-123-foo

```

This applies different OSD specs to different hosts depending on the host\_pattern key.

## Dedicated wal + db

All previous cases co-located the WALs with the DBs. It's however possible to deploy the WAL on a dedicated device as well, if it makes sense.

```

1. 20 HDDs
2. Vendor: VendorA
3. Model: SSD-123-foo
4. Size: 4TB
5.

```

```
6. 2 SSDs
7. Vendor: VendorB
8. Model: MC-55-44-ZX
9. Size: 512GB
10.
11. 2 NVMEs
12. Vendor: VendorC
13. Model: NVME-QQQQ-987
14. Size: 256GB
```

The OSD spec for this case would look like the following (using the model filter):

```
1. service_type: osd
2. service_id: osd_spec_default
3. placement:
4. host_pattern: '*'
5. data_devices:
6. model: MC-55-44-XZ
7. db_devices:
8. model: SSD-123-foo
9. wal_devices:
10. model: NVME-QQQQ-987
```

This can easily be done with other filters, like size or vendor as well.

# Troubleshooting

Sometimes there is a need to investigate why a cephadm command failed or why a specific service no longer runs properly.

As cephadm deploys daemons as containers, troubleshooting daemons is slightly different. Here are a few tools and commands to help investigating issues.

## Pausing or disabling cephadm

If something goes wrong and cephadm is doing behaving in a way you do not like, you can pause most background activity with:

```
1. ceph orch pause
```

This will stop any changes, but cephadm will still periodically check hosts to refresh its inventory of daemons and devices. You can disable cephadm completely with:

```
1. ceph orch set backend ''
2. ceph mgr module disable cephadm
```

This will disable all of the `ceph orch ...` CLI commands but the previously deployed daemon containers will still continue to exist and start as they did before.

## Checking cephadm logs

You can monitor the cephadm log in real time with:

```
1. ceph -W cephadm
```

You can see the last few messages with:

```
1. ceph log last cephadm
```

If you have enabled logging to files, you can see a cephadm log file called `ceph.cephadm.log` on monitor hosts (see [Ceph daemon logs](#)).

## Gathering log files

Use journalctl to gather the log files of all daemons:

Note

By default cephadm now stores logs in journald. This means that you will no longer find daemon logs in `/var/log/ceph/`.

To read the log file of one specific daemon, run:

```
1. cephadm logs --name <name-of-daemon>
```

Note: this only works when run on the same host where the daemon is running. To get logs of a daemon running on a different host, give the `--fsid` option:

```
1. cephadm logs --fsid <fsid> --name <name-of-daemon>
```

where the `<fsid>` corresponds to the cluster ID printed by `ceph status`.

To fetch all log files of all daemons on a given host, run:

```
1. for name in $(cephadm ls | jq -r '.[].name') ; do
2.   cephadm logs --fsid <fsid> --name "$name" > $name;
3. done
```

## Collecting systemd status

To print the state of a systemd unit, run:

```
1. systemctl status "ceph-$(cephadm shell ceph fsid)@<service name>.service";
```

To fetch all state of all daemons of a given host, run:

```
1. fsid="$(cephadm shell ceph fsid)"
2. for name in $(cephadm ls | jq -r '.[].name') ; do
3.   systemctl status "ceph-$fsid@$name.service" > $name;
4. done
```

## List all downloaded container images

To list all container images that are downloaded on a host:

Note

`Image` might also be called `ImageID`

```
1. podman ps -a --format json | jq '.[].Image'
2. "docker.io/library/centos:8"
3. "registry.opensuse.org/opensuse/leap:15.2"
```

# Manually running containers

Cephadm writes small wrappers that run a containers. Refer to `/var/lib/ceph/<cluster-fsid>/<service-name>/unit.run` for the container execution command.

## ssh errors

Error message:

```
execnet.gateway_bootstrap.HostNotFound: -F /tmp/cephadm-conf-73z09u6g -i /tmp/cephadm-identity-ky7ahp_5
1. root@10.10.1.2
2. ...
3. raise OrchestratorError(msg) from e
4. orchestrator._interface.OrchestratorError: Failed to connect to 10.10.1.2 (10.10.1.2).
5. Please make sure that the host is reachable and accepts connections using the cephadm SSH key
6. ...
```

Things users can do:

1. Ensure cephadm has an SSH identity key:

```
1. [root@mon1~]# cephadm shell -- ceph config-key get mgr/cephadm/ssh_identity_key > ~/cephadm_private_key
2. INFO:cephadm:Inferring fsid f8edc08a-7f17-11ea-8707-000c2915dd98
3. INFO:cephadm:Using recent ceph image docker.io/ceph/ceph:v15 obtained 'mgr/cephadm/ssh_identity_key'
4. [root@mon1 ~] # chmod 0600 ~/cephadm_private_key
```

If this fails, cephadm doesn't have a key. Fix this by running the following command:

```
1. [root@mon1 ~]# cephadm shell -- ceph cephadm generate-ssh-key
```

or:

```
1. [root@mon1 ~]# cat ~/cephadm_private_key | cephadm shell -- ceph cephadm set-ssk-key -i -
```

1. Ensure that the ssh config is correct:

```
1. [root@mon1 ~]# cephadm shell -- ceph cephadm get-ssh-config > config
```

2. Verify that we can connect to the host:

```
1. [root@mon1 ~]# ssh -F config -i ~/cephadm_private_key root@mon1
```

## Verifying that the Public Key is Listed in the authorized\_keys file

To verify that the public key is in the authorized\_keys file, run the following commands:

```
1. [root@mon1 ~]# cephadm shell -- ceph cephadm get-pub-key > ~/ceph.pub
2. [root@mon1 ~]# grep "`cat ~/ceph.pub`" /root/.ssh/authorized_keys
```

## Failed to infer CIDR network error

If you see this error:

```
1. ERROR: Failed to infer CIDR network for mon ip ***; pass --skip-mon-network to configure it later
```

Or this error:

```
1. Must set public_network config option or specify a CIDR network, ceph addrvec, or plain IP
```

This means that you must run a command of this form:

```
1. ceph config set mon public_network <mon_network>
```

For more detail on operations of this kind, see [Deploy additional monitors \(optional\)](#)

## Accessing the admin socket

Each Ceph daemon provides an admin socket that bypasses the MONs (See [Using the Admin Socket](#)).

To access the admin socket, first enter the daemon container on the host:

```
1. [root@mon1 ~]# cephadm enter --name <daemon-name>
2. [ceph: root@mon1 /]# ceph --admin-daemon /var/run/ceph/ceph-<daemon-name>.asok config show
```

# Cephadm Concepts

## Fully qualified domain names vs bare host names

cephadm has very minimal requirements when it comes to resolving host names etc. When cephadm initiates an ssh connection to a remote host, the host name can be resolved in four different ways:

- a custom ssh config resolving the name to an IP
- via an externally maintained `/etc/hosts`
- via explicitly providing an IP address to cephadm: `ceph orch host add <hostname> <IP>`
- automatic name resolution via DNS.

Ceph itself uses the command `hostname` to determine the name of the current host.

### Note

cephadm demands that the name of the host given via `ceph orch host add` equals the output of `hostname` on remote hosts.

Otherwise cephadm can't be sure, the host names returned by `ceph * metadata` match the hosts known to cephadm. This might result in a `CEPHADM_STRAY_HOST` warning.

When configuring new hosts, there are two **valid** ways to set the `hostname` of a host:

1. Using the bare host name. In this case:

- `hostname` returns the bare host name.
- `hostname -f` returns the FQDN.

1. Using the fully qualified domain name as the host name. In this case:

- `hostname` returns the FQDN
- `hostname -s` return the bare host name

Note that `man hostname` recommends `hostname` to return the bare host name:

The FQDN (Fully Qualified Domain Name) of the system is the name that the resolver(3) returns for the host name, such as, `ursula.example.com`. It is usually the hostname followed by the DNS domain name (the part after the first dot). You can check the FQDN using `hostname --fqdn` or the domain name using `dnsdomainname`.

1. You cannot change the FQDN with `hostname` or `dnsdomainname`.
- 2.

3. The recommended method of setting the FQDN is to make the hostname
4. be an alias for the fully qualified name using /etc/hosts, DNS, or
5. NIS. For example, if the hostname was "ursula", one might have
6. a line in /etc/hosts which reads
- 7.
8. `127.0.1.1 ursula.example.com ursula`

Which means, `man hostname` recommends `hostname` to return the bare host name. This in turn means that Ceph will return the bare host names when executing `ceph * metadata`. This in turn means cephadm also requires the bare host name when adding a host to the cluster: `ceph orch host add <bare-name>`.

## Cephadm Scheduler

Cephadm uses a declarative state to define the layout of the cluster. This state consists of a list of service specifications containing placement specifications (See [Service Specification](#)).

Cephadm constantly compares list of actually running daemons in the cluster with the desired service specifications and will either add or remove new daemons.

First, cephadm will select a list of candidate hosts. It first looks for explicit host names and will select those. In case there are no explicit hosts defined, cephadm looks for a label specification. If there is no label defined in the specification, cephadm will select hosts based on a host pattern. If there is no pattern defined, cephadm will finally select all known hosts as candidates.

Then, cephadm will consider existing daemons of this services and will try to avoid moving any daemons.

Cephadm supports the deployment of a specific amount of services. Let's consider a service specification like so:

1. `service_type: mds`
2. `service_name: myfs`
3. `placement:`
4. `count: 3`
5. `label: myfs`

This instructs cephadm to deploy three daemons on hosts labeled with `myfs` across the cluster.

Then, in case there are less than three daemons deployed on the candidate hosts, cephadm will then randomly choose hosts for deploying new daemons.

In case there are more than three daemons deployed, cephadm will remove existing daemons.

Finally, cephadm will remove daemons on hosts that are outside of the list of

candidate hosts.

However, there is a special cases that cephadm needs to consider.

In case the are fewer hosts selected by the placement specification than demanded by `count` , cephadm will only deploy on selected hosts.

# CEPHADM Developer Documentation

---

## Contents

- [Host Maintenance](#)

# Host Maintenance

All hosts that support Ceph daemons need to support maintenance activity, whether the host is physical or virtual. This means that management workflows should provide a simple and consistent way to support this operational requirement. This document defines the maintenance strategy that could be implemented in cephadm and mgr/cephadm.

## High Level Design

Placing a host into maintenance, adopts the following workflow;

1. confirm that the removal of the host does not impact data availability (the following steps will assume it is safe to proceed)
  - `orch host ok-to-stop <host>` would be used here
2. if the host has osd daemons, apply noout to the host subtree to prevent data migration from triggering during the planned maintenance slot.
3. Stop the ceph target (all daemons stop)
4. Disable the ceph target on that host, to prevent a reboot from automatically starting ceph services again)

Exiting Maintenance, is basically the reverse of the above sequence

## Admin Interaction

The ceph orch command will be extended to support maintenance.

- ```
1. ceph orch host enter-maintenance <host> [ --check ]
2. ceph orch host exit-maintenance <host>
```

### Note

In addition, the host's status should be updated to reflect whether it is in maintenance or not.

## The 'check' Option

The orch host ok-to-stop command focuses on ceph daemons (mon, osd, mds), which provides the first check. However, a ceph cluster also uses other types of daemons for monitoring, management and non-native protocol support which means the logic will need to consider service impact too. The 'check' option provides this additional layer to alert the user of service impact to secondary daemons.

The list below shows some of these additional daemons.

- mgr (not included in ok-to-stop checks)
- prometheus, grafana, alertmanager
- rgw
- haproxy
- iscsi gateways
- ganesha gateways

By using the `-check` option first, the Admin can choose whether to proceed. This workflow is obviously optional for the CLI user, but could be integrated into the UI workflow to help less experienced Administrators manage the cluster.

By adopting this two-phase approach, a UI based workflow would look something like this.

1. User selects a host to place into maintenance
  - orchestrator checks for data **and** service impact
2. If potential impact is shown, the next steps depend on the impact type
  - **data availability** : maintenance is denied, informing the user of the issue
  - **service availability** : user is provided a list of affected services and asked to confirm

## Components Impacted

---

Implementing this capability will require changes to the following;

- cephadm
  - Add maintenance subcommand with the following ‘verbs’; enter, exit, check
- mgr/cephadm
  - add methods to CephadmOrchestrator for enter/exit and check
  - data gathering would be skipped for hosts in a maintenance state
- mgr/orchestrator
  - add CLI commands to OrchestratorCli which expose the enter/exit and check interaction

## Ideas for Future Work

---

1. When a host is placed into maintenance, the time of the event could be persisted. This would allow the orchestrator layer to establish a maintenance window for the task and alert if the maintenance window has been exceeded.
2. The maintenance process could support plugins to allow other integration tasks to be initiated as part of the transition to and from maintenance. This plugin capability could support actions like;
  - alert suppression to 3rd party monitoring framework(s)
  - service level reporting, to record outage windows

# Ceph Storage Cluster

The [Ceph Storage Cluster](#) is the foundation for all Ceph deployments. Based upon RADOS, Ceph Storage Clusters consist of two types of daemons: a [Ceph OSD Daemon](#) (OSD) stores data as objects on a storage node; and a [Ceph Monitor](#) (MON) maintains a master copy of the cluster map. A Ceph Storage Cluster may contain thousands of storage nodes. A minimal system will have at least one Ceph Monitor and two Ceph OSD Daemons for data replication.

The Ceph File System, Ceph Object Storage and Ceph Block Devices read data from and write data to the Ceph Storage Cluster.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                       |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <h2>Config and Deploy</h2> <p>Ceph Storage Clusters have a few required settings, but most configuration settings have default values. A typical deployment uses a deployment tool to define a cluster and bootstrap a monitor. See <a href="#">Deployment</a> for details on <code>cephadm</code>.</p> <ul style="list-style-type: none"> <li>• Configuration           <ul style="list-style-type: none"> <li>◦ Storage devices</li> <li>◦ Configuring Ceph</li> <li>◦ Common Settings</li> <li>◦ Networks</li> <li>◦ Monitors</li> <li>◦ Authentication</li> <li>◦ OSDs</li> <li>◦ Heartbeats</li> <li>◦ Logs / Debugging</li> <li>◦ Example <code>ceph.conf</code></li> <li>◦ Running Multiple Clusters (DEPRECATED)</li> <li>◦ Network Settings</li> <li>◦ Messenger v2 protocol</li> <li>◦ Auth Settings</li> <li>◦ Monitor Settings</li> <li>◦ Looking up Monitors through DNS</li> <li>◦ Heartbeat Settings</li> <li>◦ OSD Settings</li> <li>◦ BlueStore Settings</li> <li>◦ FileStore Settings</li> <li>◦ Journal Settings</li> <li>◦ Pool, PG &amp; CRUSH Settings</li> <li>◦ Messaging Settings</li> <li>◦ General Settings</li> </ul> </li> <li>• Deployment           <ul style="list-style-type: none"> <li>◦ Stability</li> <li>◦ Deploying a new Ceph</li> </ul> </li> </ul> | <h2>Operations</h2> <p>Once you have deployed a Ceph Storage Cluster, you may begin operating your cluster.</p> <ul style="list-style-type: none"> <li>• Operations           <ul style="list-style-type: none"> <li>◦ Operating a Cluster</li> <li>◦ Health checks</li> <li>◦ Monitoring a Cluster</li> <li>◦ Monitoring OSDs and PGs</li> <li>◦ User Management</li> <li>◦ Repairing PG inconsistencies</li> <li>◦ Data Placement Overview</li> <li>◦ Pools</li> <li>◦ Erasure code</li> <li>◦ Cache Tiering</li> <li>◦ Placement Groups</li> <li>◦ Balancer</li> <li>◦ Using the pg-upmap</li> <li>◦ CRUSH Maps</li> <li>◦ Manually editing a CRUSH Map</li> <li>◦ Stretch Clusters</li> <li>◦ Configure Monitor Election Strategies</li> <li>◦ Adding/Removing OSDs</li> <li>◦ Adding/Removing Monitors</li> <li>◦ Device Management</li> <li>◦ BlueStore</li> </ul> </li> </ul> | <h2>APIs</h2> <p>Most Ceph deployments use <a href="#">Ceph Block Devices</a>, <a href="#">Ceph Object Storage</a> and/or the <a href="#">Ceph File System</a>. You may also develop applications that talk directly to the Ceph Storage Cluster.</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- Deploying a new Ceph cluster
- Converting an existing cluster to cephadm
- Upgrading Ceph
- Cephadm operations
- Cephadm monitoring
- Cephadm CLI
- Client Setup
- DriveGroups
- Troubleshooting
- Cephadm Concepts
- Cephadm Feature Planning

- Migration
- Command Reference
- The Ceph Community
- Troubleshooting Monitors
- Troubleshooting OSDs
- Troubleshooting PGs
- Logging and Debugging
- CPU Profiling
- Memory Profiling

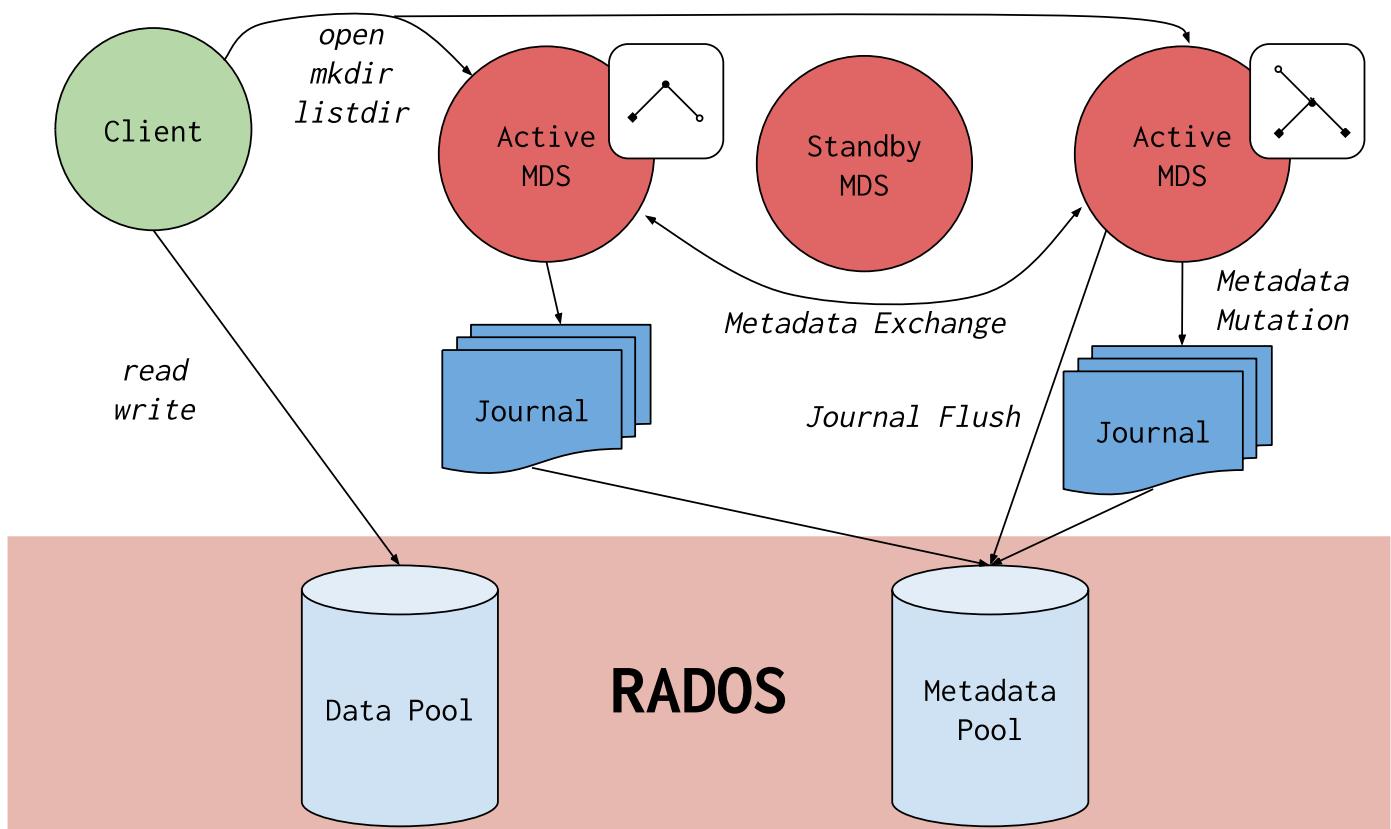
- Man Pages

# Ceph File System

The Ceph File System, or **CephFS**, is a POSIX-compliant file system built on top of Ceph's distributed object store, **RADOS**. CephFS endeavors to provide a state-of-the-art, multi-use, highly available, and performant file store for a variety of applications, including traditional use-cases like shared home directories, HPC scratch space, and distributed workflow shared storage.

CephFS achieves these goals through the use of some novel architectural choices. Notably, file metadata is stored in a separate RADOS pool from file data and served via a resizable cluster of *Metadata Servers*, or **MDS**, which may scale to support higher throughput metadata workloads. Clients of the file system have direct access to RADOS for reading and writing file data blocks. For this reason, workloads may linearly scale with the size of the underlying RADOS object store; that is, there is no gateway or broker mediating data I/O for clients.

Access to data is coordinated through the cluster of MDS which serve as authorities for the state of the distributed metadata cache cooperatively maintained by clients and MDS. Mutations to metadata are aggregated by each MDS into a series of efficient writes to a journal on RADOS; no metadata state is stored locally by the MDS. This model allows for coherent and rapid collaboration between clients within the context of a POSIX file system.



CephFS is the subject of numerous academic papers for its novel designs and

contributions to file system research. It is the oldest storage interface in Ceph and was once the primary use-case for RADOS. Now it is joined by two other storage interfaces to form a modern unified storage system: RBD (Ceph Block Devices) and RGW (Ceph Object Storage Gateway).

## Getting Started with CephFS

---

For most deployments of Ceph, setting up a CephFS file system is as simple as:

```
1. ceph fs volume create <fs name>
```

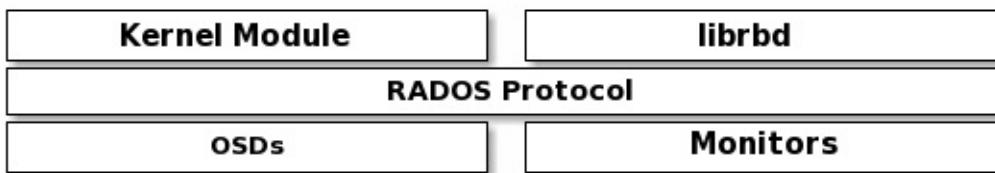
The Ceph [Orchestrator](#) will automatically create and configure MDS for your file system if the back-end deployment technology supports it (see [Orchestrator deployment table](#)). Otherwise, please [deploy MDS manually as needed](#).

Finally, to mount CephFS on your client nodes, see [Mount CephFS: Prerequisites](#) page. Additionally, a command-line shell utility is available for interactive access or scripting via the [cephfs-shell](#).

# Ceph Block Device

A block is a sequence of bytes (often 512). Block-based storage interfaces are a mature and common way to store data on media including HDDs, SSDs, CDS, floppy disks, and even tape. The ubiquity of block device interfaces is a perfect fit for interacting with mass data storage including Ceph.

Ceph block devices are thin-provisioned, resizable, and store data striped over multiple OSDs. Ceph block devices leverage RADOS capabilities including snapshotting, replication and strong consistency. Ceph block storage clients communicate with Ceph clusters through kernel modules or the `librbd` library.



## Note

Kernel modules can use Linux page caching. For `librbd`-based applications, Ceph supports [RBD Caching](#).

Ceph's block devices deliver high performance with vast scalability to [kernel modules](#), or to KVMs such as [QEMU](#), and cloud-based computing systems like [OpenStack](#) and [CloudStack](#) that rely on libvirt and QEMU to integrate with Ceph block devices. You can use the same cluster to operate the [Ceph RADOS Gateway](#), the [Ceph File System](#), and Ceph block devices simultaneously.

## Important

To use Ceph Block Devices, you must have access to a running Ceph cluster.

- [Basic Commands](#)
- [Operations](#)
  - [Snapshots](#)
  - [Exclusive Locking](#)
  - [Mirroring](#)
  - [Live-Migration](#)
  - [Persistent Cache](#)
  - [Config Settings \(librbd\)](#)
  - [RBD Replay](#)

- Integrations

- Kernel Modules
- QEMU
- libvirt
- Kubernetes
- OpenStack
- CloudStack
- LIO iSCSI Gateway

- Manpages

- rbd
- rbd-fuse
- rbd-nbd
- rbd-ggate
- rbd-map
- ceph-rbdnamer
- rbd-replay-prep
- rbd-replay
- rbd-replay-many

# Snapshots

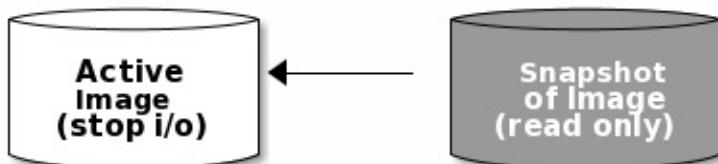
A snapshot is a read-only logical copy of an image at a particular point in time: a checkpoint. One of the advanced features of Ceph block devices is that you can create snapshots of images to retain point-in-time state history. Ceph also supports snapshot layering, which allows you to clone images (e.g., a VM image) quickly and easily. Ceph block device snapshots are managed using the `rbd` command and multiple higher level interfaces, including `QEMU`, `libvirt`, `OpenStack` and `CloudStack`.

## Important

To use RBD snapshots, you must have a running Ceph cluster.

## Note

Because RBD does not know about any filesystem within an image (volume), snapshots are not crash-consistent unless they are coordinated within the mounting (attaching) operating system. We therefore recommend that you pause or stop I/O before taking a snapshot. If the volume contains a filesystem, it must be in an internally consistent state before taking a snapshot. Snapshots taken at inconsistent points may need a fsck pass before subsequent mounting. To stop I/O you can use `fsfreeze` command. See `fsfreeze(8)` man page for more details. For virtual machines, `qemu-guest-agent` can be used to automatically freeze file systems when creating a snapshot.



# Ceph Notes

When `cephx` authentication is enabled (it is by default), you must specify a user name or ID and a path to the keyring containing the corresponding key. See [User Management](#) for details. You may also set the `CEPH_ARGS` environment variable to avoid re-entry of these parameters.

1. `rbd --id {user-ID} --keyring=/path/to/secret [commands]`
2. `rbd --name {username} --keyring=/path/to/secret [commands]`

For example:

1. `rbd --id admin --keyring=/etc/ceph/ceph.keyring [commands]`
2. `rbd --name client.admin --keyring=/etc/ceph/ceph.keyring [commands]`

## Tip

Add the user and secret to the `CEPH_ARGS` environment variable so that you don't need to enter them each time.

# Snapshot Basics

The following procedures demonstrate how to create, list, and remove snapshots using the `rbd` command.

## Create Snapshot

To create a snapshot with `rbd`, specify the `snap create` option, the pool name and the image name.

```
1. rbd snap create {pool-name}/{image-name}@{snap-name}
```

For example:

```
1. rbd snap create rbd/foo@snapname
```

## List Snapshots

To list snapshots of an image, specify the pool name and the image name.

```
1. rbd snap ls {pool-name}/{image-name}
```

For example:

```
1. rbd snap ls rbd/foo
```

## Rollback Snapshot

To rollback to a snapshot with `rbd`, specify the `snap rollback` option, the pool name, the image name and the snap name.

```
1. rbd snap rollback {pool-name}/{image-name}@{snap-name}
```

For example:

```
1. rbd snap rollback rbd/foo@snapname
```

## Note

Rolling back an image to a snapshot means overwriting the current version of the image with data from a snapshot. The time it takes to execute a rollback increases with the size of the image. It is **faster to clone** from a snapshot **than to rollback** an image to a snapshot, and is the preferred method of returning to a pre-existing state.

## Delete a Snapshot

To delete a snapshot with `rbd`, specify the `snap rm` subcommand, the pool name, the image name and the snap name.

```
1. rbd snap rm {pool-name}/{image-name}@{snap-name}
```

For example:

```
1. rbd snap rm rbd/foo@snapname
```

## Note

Ceph OSDs delete data asynchronously, so deleting a snapshot doesn't immediately free up the underlying OSDs' capacity.

## Purge Snapshots

To delete all snapshots for an image with `rbd`, specify the `snap purge` subcommand and the image name.

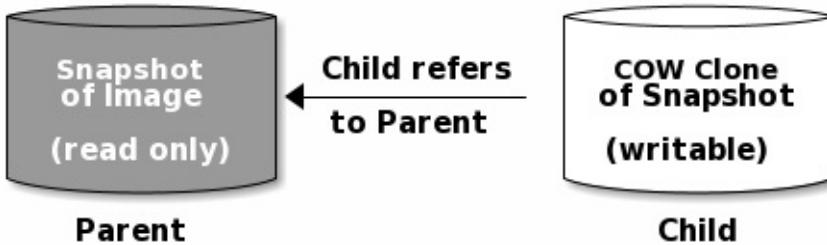
```
1. rbd snap purge {pool-name}/{image-name}
```

For example:

```
1. rbd snap purge rbd/foo
```

## Layering

Ceph supports the ability to create many copy-on-write (COW) clones of a block device snapshot. Snapshot layering enables Ceph block device clients to create images very quickly. For example, you might create a block device image with a Linux VM written to it; then, snapshot the image, protect the snapshot, and create as many copy-on-write clones as you like. A snapshot is read-only, so cloning a snapshot simplifies semantics-making it possible to create clones rapidly.



#### Note

The terms “parent” and “child” refer to a Ceph block device snapshot (parent), and the corresponding image cloned from the snapshot (child). These terms are important for the command line usage below.

Each cloned image (child) stores a reference to its parent image, which enables the cloned image to open the parent snapshot and read it.

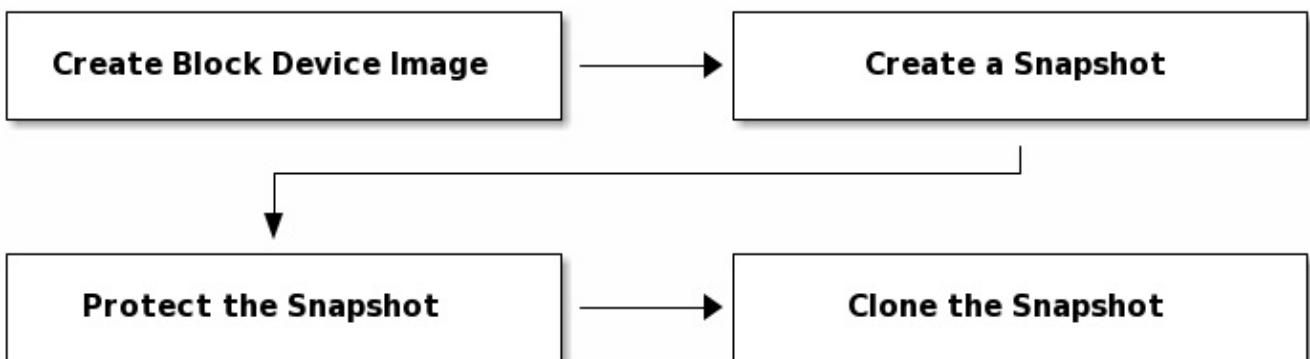
A COW clone of a snapshot behaves exactly like any other Ceph block device image. You can read to, write from, clone, and resize cloned images. There are no special restrictions with cloned images. However, the copy-on-write clone of a snapshot depends on the snapshot, so you **MUST** protect the snapshot before you clone it. The following diagram depicts the process.

#### Note

Ceph only supports cloning of RBD format 2 images (i.e., created with `rbd create --image-format 2`). The kernel client supports cloned images beginning with the 3.10 release.

## Getting Started with Layering

Ceph block device layering is a simple process. You must have an image. You must create a snapshot of the image. You must protect the snapshot. Once you have performed these steps, you can begin cloning the snapshot.



The cloned image has a reference to the parent snapshot, and includes the pool ID, image ID and snapshot ID. The inclusion of the pool ID means that you may clone snapshots from one pool to images in another pool.

1. **Image Template:** A common use case for block device layering is to create a master image and a snapshot that serves as a template for clones. For example, a user may create an image for a Linux distribution (e.g., Ubuntu 12.04), and create a snapshot for it. Periodically, the user may update the image and create a new snapshot (e.g., `sudo apt-get update` , `sudo apt-get upgrade` , `sudo apt-get dist-upgrade` followed by `rbd snap create` ). As the image matures, the user can clone any one of the snapshots.
2. **Extended Template:** A more advanced use case includes extending a template image that provides more information than a base image. For example, a user may clone an image (e.g., a VM template) and install other software (e.g., a database, a content management system, an analytics system, etc.) and then snapshot the extended image, which itself may be updated just like the base image.
3. **Template Pool:** One way to use block device layering is to create a pool that contains master images that act as templates, and snapshots of those templates. You may then extend read-only privileges to users so that they may clone the snapshots without the ability to write or execute within the pool.
4. **Image Migration/Recovery:** One way to use block device layering is to migrate or recover data from one pool into another pool.

## Protecting a Snapshot

Clones access the parent snapshots. All clones would break if a user inadvertently deleted the parent snapshot. To prevent data loss, you **MUST** protect the snapshot before you can clone it.

```
1. rbd snap protect {pool-name}/{image-name}@{snapshot-name}
```

For example:

```
1. rbd snap protect rbd/my-image@my-snapshot
```

### Note

You cannot delete a protected snapshot.

## Cloning a Snapshot

To clone a snapshot, specify you need to specify the parent pool, image and snapshot; and, the child pool and image name. You must protect the snapshot before you can clone it.

```
1. rbd clone {pool-name}/{parent-image}@{snap-name} {pool-name}/{child-image-name}
```

For example:

```
1. rbd clone rbd/my-image@my-snapshot rbd/new-image
```

#### Note

You may clone a snapshot from one pool to an image in another pool. For example, you may maintain read-only images and snapshots as templates in one pool, and writeable clones in another pool.

## Unprotecting a Snapshot

Before you can delete a snapshot, you must unprotect it first. Additionally, you may *NOT* delete snapshots that have references from clones. You must flatten each clone of a snapshot, before you can delete the snapshot.

```
1. rbd snap unprotect {pool-name}/{image-name}@{snapshot-name}
```

For example:

```
1. rbd snap unprotect rbd/my-image@my-snapshot
```

## Listing Children of a Snapshot

To list the children of a snapshot, execute the following:

```
1. rbd children {pool-name}/{image-name}@{snapshot-name}
```

For example:

```
1. rbd children rbd/my-image@my-snapshot
```

## Flattening a Cloned Image

Cloned images retain a reference to the parent snapshot. When you remove the reference from the child clone to the parent snapshot, you effectively “flatten” the image by copying the information from the snapshot to the clone. The time it takes to flatten a clone increases with the size of the snapshot. To delete a snapshot, you must flatten the child images first.

```
1. rbd flatten {pool-name}/{image-name}
```

For example:

```
1. rbd flatten rbd/new-image
```

Note

Since a flattened image contains all the information from the snapshot, a flattened image will take up more storage space than a layered clone.

# RBD Exclusive Locks

Exclusive locks are a mechanism designed to prevent multiple processes from accessing the same Rados Block Device (RBD) in an uncoordinated fashion. Exclusive locks are heavily used in virtualization (where they prevent VMs from clobbering each others' writes), and also in RBD mirroring (where they are a prerequisite for journaling).

Exclusive locks are enabled on newly created images by default, unless overridden via the `rbd_default_features` configuration option or the `--image-feature` flag for `rbd create`.

In order to ensure proper exclusive locking operations, any client using an RBD image whose `exclusive-lock` feature is enabled should be using a CephX identity whose capabilities include `profile rbd`.

Exclusive locking is mostly transparent to the user.

1. Whenever any `librbd` client process or kernel RBD client starts using an RBD image on which exclusive locking has been enabled, it obtains an exclusive lock on the image before the first write.
2. Whenever any such client process gracefully terminates, it automatically relinquishes the lock.
3. This subsequently enables another process to acquire the lock, and write to the image.

Note that it is perfectly possible for two or more concurrently running processes to merely open the image, and also to read from it. The client acquires the exclusive lock only when attempting to write to the image. To disable transparent lock transitions between multiple clients, it needs to acquire the lock specifically with `RBD_LOCK_MODE_EXCLUSIVE`.

## Blacklisting

Sometimes, a client process (or, in case of a krbd client, a client node's kernel thread) that previously held an exclusive lock on an image does not terminate gracefully, but dies abruptly. This may be due to having received a `KILL` or `ABRT` signal, for example, or a hard reboot or power failure of the client node. In that case, the exclusive lock is never gracefully released. Thus, when a new process starts and attempts to use the device, it needs a way to break the previously held exclusive lock.

However, a process (or kernel thread) may also hang, or merely lose network connectivity to the Ceph cluster for some amount of time. In that case, simply breaking the lock would be potentially catastrophic: the hung process or connectivity issue may resolve itself, and the old process may then compete with one that has

started in the interim, accessing RBD data in an uncoordinated and destructive manner.

Thus, in the event that a lock cannot be acquired in the standard graceful manner, the overtaking process not only breaks the lock, but also blocklists the previous lock holder. This is negotiated between the new client process and the Ceph Mon: upon receiving the blocklist request,

- the Mon instructs the relevant OSDs to no longer serve requests from the old client process;
- once the associated OSD map update is complete, the Mon grants the lock to the new client;
- once the new client has acquired the lock, it can commence writing to the image.

Blocklisting is thus a form of storage-level resource [fencing](#)).

In order for blocklisting to work, the client must have the `osd blocklist` capability.

This capability is included in the `profile rbd` capability profile, which should generally be set on all Ceph [client identities](#) using RBD.

# RBD Mirroring

RBD images can be asynchronously mirrored between two Ceph clusters. This capability is available in two modes:

- **Journal-based:** This mode uses the RBD journaling image feature to ensure point-in-time, crash-consistent replication between clusters. Every write to the RBD image is first recorded to the associated journal before modifying the actual image. The remote cluster will read from this associated journal and replay the updates to its local copy of the image. Since each write to the RBD image will result in two writes to the Ceph cluster, expect write latencies to nearly double while using the RBD journaling image feature.
- **Snapshot-based:** This mode uses periodically scheduled or manually created RBD image mirror-snapshots to replicate crash-consistent RBD images between clusters. The remote cluster will determine any data or metadata updates between two mirror-snapshots and copy the deltas to its local copy of the image. With the help of the RBD `fast-diff` image feature, updated data blocks can be quickly determined without the need to scan the full RBD image. Since this mode is not as fine-grained as journaling, the complete delta between two snapshots will need to be synced prior to use during a failover scenario. Any partially applied set of deltas will be rolled back at moment of failover.

## Note

journal-based mirroring requires the Ceph Jewel release or later; snapshot-based mirroring requires the Ceph Octopus release or later.

Mirroring is configured on a per-pool basis within peer clusters and can be configured on a specific subset of images within the pool. You can also mirror all images within a given pool when using journal-based mirroring. Mirroring is configured using the `rbd` command. The `rbd-mirror` daemon is responsible for pulling image updates from the remote peer cluster and applying them to the image within the local cluster.

Depending on the desired needs for replication, RBD mirroring can be configured for either one- or two-way replication:

- **One-way Replication:** When data is only mirrored from a primary cluster to a secondary cluster, the `rbd-mirror` daemon runs only on the secondary cluster.
- **Two-way Replication:** When data is mirrored from primary images on one cluster to non-primary images on another cluster (and vice-versa), the `rbd-mirror` daemon runs on both clusters.

## Important

Each instance of the `rbd-mirror` daemon must be able to connect to both the local and

remote Ceph clusters simultaneously (i.e. all monitor and OSD hosts). Additionally, the network must have sufficient bandwidth between the two data centers to handle mirroring workload.

## Pool Configuration

The following procedures demonstrate how to perform the basic administrative tasks to configure mirroring using the `rbd` command. Mirroring is configured on a per-pool basis.

These pool configuration steps should be performed on both peer clusters. These procedures assume that both clusters, named “site-a” and “site-b”, are accessible from a single host for clarity.

See the `rbd` manpage for additional details of how to connect to different Ceph clusters.

### Note

The cluster name in the following examples corresponds to a Ceph configuration file of the same name (e.g. `/etc/ceph/site-b.conf`). See the `ceph-conf` documentation for how to configure multiple clusters. Note that `rbd-mirror` does **not** require the source and destination clusters to have unique internal names; both can and should call themselves `ceph`. The config files that `rbd-mirror` needs for local and remote clusters can be named arbitrarily, and containerizing the daemon is one strategy for maintaining them outside of `/etc/ceph` to avoid confusion.

## Enable Mirroring

To enable mirroring on a pool with `rbd`, issue the `mirror pool enable` subcommand with the pool name, and the mirroring mode:

```
1. rbd mirror pool enable {pool-name} {mode}
```

The mirroring mode can either be `image` or `pool`:

- **image**: When configured in `image` mode, mirroring must **explicitly enabled** on each image.
- **pool** (default): When configured in `pool` mode, all images in the pool with the journaling feature enabled are mirrored.

For example:

```
1. $ rbd --cluster site-a mirror pool enable image-pool image
2. $ rbd --cluster site-b mirror pool enable image-pool image
```

## Disable Mirroring

To disable mirroring on a pool with `rbd`, specify the `mirror pool disable` command and the pool name:

```
1. rbd mirror pool disable {pool-name}
```

When mirroring is disabled on a pool in this way, mirroring will also be disabled on any images (within the pool) for which mirroring was enabled explicitly.

For example:

```
1. $ rbd --cluster site-a mirror pool disable image-pool
2. $ rbd --cluster site-b mirror pool disable image-pool
```

## Bootstrap Peers

In order for the `rbd-mirror` daemon to discover its peer cluster, the peer must be registered and a user account must be created. This process can be automated with `rbd` and the `mirror pool peer bootstrap create` and `mirror pool peer bootstrap import` commands.

To manually create a new bootstrap token with `rbd`, issue the `mirror pool peer bootstrap create` subcommand, a pool name, and an optional friendly site name to describe the local cluster:

```
1. rbd mirror pool peer bootstrap create [--site-name {local-site-name}] {pool-name}
```

The output of `mirror pool peer bootstrap create` will be a token that should be provided to the `mirror pool peer bootstrap import` command. For example, on site-a:

```
1. $ rbd --cluster site-a mirror pool peer bootstrap create --site-name site-a image-pool
2.eyJmc2lkIjoi0WY1MjgyZGItYjg5OS00NTk2LTgw0TgtMzIwYzM5NmYzIiwiY2xpZW50X2lkIjoicmJkLW1pcnJvciiwZWVyiawia2V5Ijoi
```

To manually import the bootstrap token created by another cluster with `rbd`, specify the `mirror pool peer bootstrap import` command, the pool name, a file path to the created token (or `'-'` to read from standard input), along with an optional friendly site name to describe the local cluster and a mirroring direction (defaults to `rx-tx` for bidirectional mirroring, but can also be set to `rx-only` for unidirectional mirroring):

```
1. rbd mirror pool peer bootstrap import [--site-name {local-site-name}] [--direction {rx-only or rx-tx}] {pool-name} {token-path}
```

For example, on site-b:

```
1. $ cat <<EOF > token
```

```

2. eyJmc2lkIjoiOWY1MjgyZGItYjg50S00NTk2LTgwOTgtMzIwYzFmYzM5NmYzIiwiY2xpZW50X2lkIjoicmJkLW1pcnJvciiwZWVyiia2V5IjoiC
3. EOF
4. $ rbd --cluster site-b mirror pool peer bootstrap import --site-name site-b image-pool token

```

## Add Cluster Peer Manually

Cluster peers can be specified manually if desired or if the above bootstrap commands are not available with the currently installed Ceph release.

The remote `rbd-mirror` daemon will need access to the local cluster to perform mirroring. A new local Ceph user should be created for the remote daemon to use. To [create a Ceph user](#), with `ceph` specify the `auth get-or-create` command, user name, monitor caps, and OSD caps:

```
1. ceph auth get-or-create client.rbd-mirror-peer mon 'profile rbd' osd 'profile rbd'
```

The resulting keyring should be copied to the other cluster's `rbd-mirror` daemon hosts if not using the Ceph monitor `config-key` store described below.

To manually add a mirroring peer Ceph cluster with `rbd`, specify the `mirror pool peer add` command, the pool name, and a cluster specification:

```
1. rbd mirror pool peer add {pool-name} {client-name}@{cluster-name}
```

For example:

```

1. $ rbd --cluster site-a mirror pool peer add image-pool client.rbd-mirror-peer@site-b
2. $ rbd --cluster site-b mirror pool peer add image-pool client.rbd-mirror-peer@site-a

```

By default, the `rbd-mirror` daemon needs to have access to a Ceph configuration file located at `/etc/ceph/{cluster-name}.conf` that provides the addresses of the peer cluster's monitors, in addition to a keyring for `{client-name}` located in the default or configured keyring search paths (e.g. `/etc/ceph/{cluster-name}.{client-name}.keyring`).

Alternatively, the peer cluster's monitor and/or client key can be securely stored within the local Ceph monitor `config-key` store. To specify the peer cluster connection attributes when adding a mirroring peer, use the `--remote-mon-host` and `--remote-key-file` optionals. For example:

```

1. $ cat <<EOF > remote-key-file
2. AQEuZdbMMoBChAACj++/XUxNOLFawdtTREEs===
3. EOF
$ rbd --cluster site-a mirror pool peer add image-pool client.rbd-mirror-peer@site-b --remote-mon-host
4. 192.168.1.1,192.168.1.2 --remote-key-file remote-key-file
5. $ rbd --cluster site-a mirror pool info image-pool --all
6. Mode: pool
7. Peers:

```

| 8. UUID                                | NAME   | CLIENT                 | MON_HOST                | KEY |
|----------------------------------------|--------|------------------------|-------------------------|-----|
| 587b08db-3d33-4f32-8af8-421e77abb081   | site-b | client.rbd-mirror-peer | 192.168.1.1,192.168.1.2 |     |
| AQAvZdbMMoBChAAcj++/XUxNOLFaWdtTRESw== |        |                        |                         |     |

## Remove Cluster Peer

To remove a mirroring peer Ceph cluster with `rbd`, specify the `mirror pool peer remove` command, the pool name, and the peer UUID (available from the `rbd mirror pool info` command):

```
1. rbd mirror pool peer remove {pool-name} {peer-uuid}
```

For example:

```
1. $ rbd --cluster site-a mirror pool peer remove image-pool 55672766-c02b-4729-8567-f13a66893445
2. $ rbd --cluster site-b mirror pool peer remove image-pool 60c0e299-b38f-4234-91f6-eed0a367be08
```

## Data Pools

When creating images in the destination cluster, `rbd-mirror` selects a data pool as follows:

1. If the destination cluster has a default data pool configured (with the `rbd_default_data_pool` configuration option), it will be used.
2. Otherwise, if the source image uses a separate data pool, and a pool with the same name exists on the destination cluster, that pool will be used.
3. If neither of the above is true, no data pool will be set.

## Image Configuration

Unlike pool configuration, image configuration only needs to be performed against a single mirroring peer Ceph cluster.

Mirrored RBD images are designated as either primary or non-primary. This is a property of the image and not the pool. Images that are designated as non-primary cannot be modified.

Images are automatically promoted to primary when mirroring is first enabled on an image (either implicitly if the pool mirror mode was `pool` and the image has the journaling image feature enabled, or explicitly enabled by the `rbd` command if the pool mirror mode was `image`).

## Enable Image Mirroring

If mirroring is configured in `image` mode for the image's pool, then it is necessary to explicitly enable mirroring for each image within the pool. To enable mirroring for a specific image with `rbd`, specify the `mirror image enable` command along with the pool, image name, and mode:

```
1. rbd mirror image enable {pool-name}/{image-name} {mode}
```

The `mirror image` mode can either be `journal` or `snapshot`:

- **journal** (default): When configured in `journal` mode, mirroring will utilize the RBD journaling image feature to replicate the image contents. If the RBD journaling image feature is not yet enabled on the image, it will be automatically enabled.
- **snapshot**: When configured in `snapshot` mode, mirroring will utilize RBD image mirror-snapshots to replicate the image contents. Once enabled, an initial mirror-snapshot will automatically be created. Additional RBD image `mirrorsnapshots` can be created by the `rbd` command.

For example:

```
1. $ rbd --cluster site-a mirror image enable image-pool/image-1 snapshot
2. $ rbd --cluster site-a mirror image enable image-pool/image-2 journal
```

## Enable Image Journaling Feature

RBD journal-based mirroring uses the RBD image journaling feature to ensure that the replicated image always remains crash-consistent. When using the `image` mirroring mode, the journaling feature will be automatically enabled when mirroring is enabled on the image. When using the `pool` mirroring mode, before an image can be mirrored to a peer cluster, the RBD image journaling feature must be enabled. The feature can be enabled at image creation time by providing the `--image-feature exclusive-lock,journaling` option to the `rbd` command.

Alternatively, the journaling feature can be dynamically enabled on pre-existing RBD images. To enable journaling with `rbd`, specify the `feature enable` command, the pool and image name, and the feature name:

```
1. rbd feature enable {pool-name}/{image-name} {feature-name}
```

For example:

```
1. $ rbd --cluster site-a feature enable image-pool/image-1 journaling
```

### Note

The journaling feature is dependent on the exclusive-lock feature. If the exclusive-lock feature is not already enabled, it should be enabled prior to enabling the journaling feature.

#### Tip

You can enable journaling on all new images by default by adding `rbd default features = 125` to your Ceph configuration file.

#### Tip

`rbd-mirror` tunables are set by default to values suitable for mirroring an entire pool. When using `rbd-mirror` to migrate single volumes between clusters you may achieve substantial performance gains by setting `rbd_mirror_journal_max_fetch_bytes=33554432` and `rbd_journal_max_payload_bytes=8388608` within the `[client]` config section of the local or centralized configuration. Note that these settings may allow `rbd-mirror` to present a substantial write workload to the destination cluster: monitor cluster performance closely during migrations and test carefully before running multiple migrations in parallel.

## Create Image Mirror-Snapshots

When using snapshot-based mirroring, mirror-snapshots will need to be created whenever it is desired to mirror the changed contents of the RBD image. To create a mirror-snapshot manually with `rbd`, specify the `mirror image snapshot` command along with the pool and image name:

```
1. rbd mirror image snapshot {pool-name}/{image-name}
```

For example:

```
1. $ rbd --cluster site-a mirror image snapshot image-pool/image-1
```

By default only `3` mirror-snapshots will be created per-image. The most recent mirror-snapshot is automatically pruned if the limit is reached. The limit can be overridden via the `rbd_mirroring_max_mirroring_snapshots` configuration option if required. Additionally, mirror-snapshots are automatically deleted when the image is removed or when mirroring is disabled.

Mirror-snapshots can also be automatically created on a periodic basis if mirror-snapshot schedules are defined. The mirror-snapshot can be scheduled globally, per-pool, or per-image levels. Multiple mirror-snapshot schedules can be defined at any level, but only the most-specific snapshot schedules that match an individual mirrored image will run.

To create a mirror-snapshot schedule with `rbd`, specify the `mirror snapshot schedule add` command along with an optional pool or image name; interval; and optional start time:

```
1. rbd mirror snapshot schedule add [--pool {pool-name}] [--image {image-name}] {interval} [{start-time}]
```

The `interval` can be specified in days, hours, or minutes using `d`, `h`, `m` suffix respectively. The optional `start-time` can be specified using the ISO 8601 time format. For example:

```
1. $ rbd --cluster site-a mirror snapshot schedule add --pool image-pool 24h 14:00:00-05:00
2. $ rbd --cluster site-a mirror snapshot schedule add --pool image-pool --image image1 6h
```

To remove a mirror-snapshot schedules with `rbd`, specify the `mirror snapshot schedule remove` command with options that match the corresponding `add` schedule command.

To list all snapshot schedules for a specific level (global, pool, or image) with `rbd`, specify the `mirror snapshot schedule ls` command along with an optional pool or image name. Additionally, the `--recursive` option can be specified to list all schedules at the specified level and below. For example:

```
1. $ rbd --cluster site-a mirror schedule ls --pool image-pool --recursive
2. POOL      NAMESPACE IMAGE  SCHEDULE
3. image-pool -      -      every 1d starting at 14:00:00-05:00
4. image-pool      image1 every 6h
```

To view the status for when the next snapshots will be created for snapshot-based mirroring RBD images with `rbd`, specify the `mirror snapshot schedule status` command along with an optional pool or image name:

```
1. rbd mirror snapshot schedule status [--pool {pool-name}] [--image {image-name}]
```

For example:

```
1. $ rbd --cluster site-a mirror schedule status
2. SCHEDULE TIME      IMAGE
3. 2020-02-26 18:00:00 image-pool/image1
```

## Disable Image Mirroring

To disable mirroring for a specific image with `rbd`, specify the `mirror image disable` command along with the pool and image name:

```
1. rbd mirror image disable {pool-name}/{image-name}
```

For example:

```
1. $ rbd --cluster site-a mirror image disable image-pool/image-1
```

## Image Promotion and Demotion

In a failover scenario where the primary designation needs to be moved to the image in the peer Ceph cluster, access to the primary image should be stopped (e.g. power down the VM or remove the associated drive from a VM), demote the current primary image, promote the new primary image, and resume access to the image on the alternate cluster.

### Note

RBD only provides the necessary tools to facilitate an orderly failover of an image. An external mechanism is required to coordinate the full failover process (e.g. closing the image before demotion).

To demote a specific image to non-primary with `rbd`, specify the `mirror image demote` command along with the pool and image name:

```
1. rbd mirror image demote {pool-name}/{image-name}
```

For example:

```
1. $ rbd --cluster site-a mirror image demote image-pool/image-1
```

To demote all primary images within a pool to non-primary with `rbd`, specify the `mirror pool demote` command along with the pool name:

```
1. rbd mirror pool demote {pool-name}
```

For example:

```
1. $ rbd --cluster site-a mirror pool demote image-pool
```

To promote a specific image to primary with `rbd`, specify the `mirror image promote` command along with the pool and image name:

```
1. rbd mirror image promote [--force] {pool-name}/{image-name}
```

For example:

```
1. $ rbd --cluster site-b mirror image promote image-pool/image-1
```

To promote all non-primary images within a pool to primary with `rbd`, specify the `mirror pool promote` command along with the pool name:

```
1. rbd mirror pool promote [--force] {pool-name}
```

For example:

```
1. $ rbd --cluster site-a mirror pool promote image-pool
```

#### Tip

Since the primary / non-primary status is per-image, it is possible to have two clusters split the IO load and stage failover / failback.

#### Note

Promotion can be forced using the `--force` option. Forced promotion is needed when the demotion cannot be propagated to the peer Ceph cluster (e.g. Ceph cluster failure, communication outage). This will result in a split-brain scenario between the two peers and the image will no longer be in-sync until a [force resync command](#) is issued.

## Force Image Resync

If a split-brain event is detected by the `rbd-mirror` daemon, it will not attempt to mirror the affected image until corrected. To resume mirroring for an image, first [demote the image](#) determined to be out-of-date and then request a resync to the primary image. To request an image resync with `rbd`, specify the `mirror image resync` command along with the pool and image name:

```
1. rbd mirror image resync {pool-name}/{image-name}
```

For example:

```
1. $ rbd mirror image resync image-pool/image-1
```

#### Note

The `rbd` command only flags the image as requiring a resync. The local cluster's `rbd-mirror` daemon process is responsible for performing the resync asynchronously.

## Mirror Status

The peer cluster replication status is stored for every primary mirrored image. This status can be retrieved using the `mirror image status` and `mirror pool status` commands.

To request the mirror image status with `rbd`, specify the `mirror image status` command along with the pool and image name:

```
1. rbd mirror image status {pool-name}/{image-name}
```

For example:

```
1. $ rbd mirror image status image-pool/image-1
```

To request the mirror pool summary status with `rbd`, specify the `mirror pool status` command along with the pool name:

```
1. rbd mirror pool status {pool-name}
```

For example:

```
1. $ rbd mirror pool status image-pool
```

#### Note

Adding `--verbose` option to the `mirror pool status` command will additionally output status details for every mirroring image in the pool.

## rbd-mirror Daemon

The two `rbd-mirror` daemons are responsible for watching image journals on the remote, peer cluster and replaying the journal events against the local cluster. The RBD image journaling feature records all modifications to the image in the order they occur. This ensures that a crash-consistent mirror of the remote image is available locally.

The `rbd-mirror` daemon is available within the optional `rbd-mirror` distribution package.

#### Important

Each `rbd-mirror` daemon requires the ability to connect to both clusters simultaneously.

#### Warning

Pre-Luminous releases: only run a single `rbd-mirror` daemon per Ceph cluster.

Each `rbd-mirror` daemon should use a unique Ceph user ID. To [create a Ceph user](#), with `ceph` specify the `auth get-or-create` command, user name, monitor caps, and OSD caps:

```
1. ceph auth get-or-create client.rbd-mirror.{unique id} mon 'profile rbd-mirror' osd 'profile rbd'
```

The `rbd-mirror` daemon can be managed by `systemd` by specifying the user ID as the daemon instance:

```
1. systemctl enable ceph-rbd-mirror@rbd-mirror.{unique id}
```

The `rbd-mirror` can also be run in foreground by `rbd-mirror` command:

```
1. rbd-mirror -f --log-file={log_path}
```

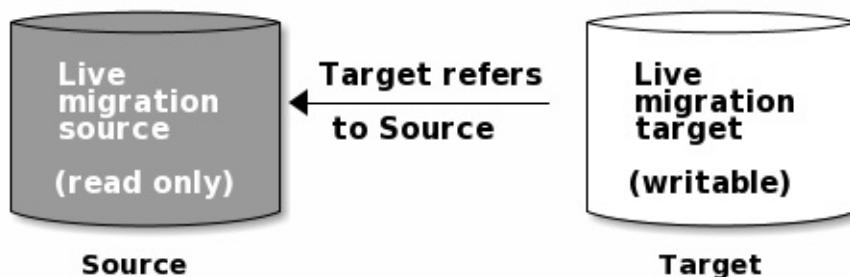
# Image Live-Migration

RBD images can be live-migrated between different pools within the same cluster or between different image formats and layouts. When started, the source image will be deep-copied to the destination image, pulling all snapshot history and optionally preserving any link to the source image's parent to preserve sparseness.

This copy process can safely run in the background while the new target image is in use. There is currently a requirement to temporarily stop using the source image before preparing a migration. This helps to ensure that the client using the image is updated to point to the new target image.

## Note

Image live-migration requires the Ceph Nautilus release or later. The `krbd` kernel module does not support live-migration at this time.



The live-migration process is comprised of three steps:

- Prepare Migration:** The initial step creates the new target image and cross-links the source and target images. Similar to [layered images](#), attempts to read uninitialized extents within the target image will internally redirect the read to the source image, and writes to uninitialized extents within the target will internally deep-copy the overlapping source image block to the target image.
- Execute Migration:** This is a background operation that deep-copies all initialized blocks from the source image to the target. This step can be run while clients are actively using the new target image.
- Finish Migration:** Once the background migration process has completed, the migration can be committed or aborted. Committing the migration will remove the cross-links between the source and target images, and will remove the source image. Aborting the migration will remove the cross-links, and will remove the target image.

## Prepare Migration

The live-migration process is initiated by running the rbd migration prepare command, providing the source and target images:

```
1. $ rbd migration prepare migration_source [migration_target]
```

The rbd migration prepare command accepts all the same layout optionals as the rbd create command, which allows changes to the immutable image on-disk layout. The migration\_target can be skipped if the goal is only to change the on-disk layout, keeping the original image name.

All clients using the source image must be stopped prior to preparing a live-migration. The prepare step will fail if it finds any running clients with the image open in read/write mode. Once the prepare step is complete, the clients can be restarted using the new target image name. Attempting to restart the clients using the source image name will result in failure.

The rbd status command will show the current state of the live-migration:

```
1. $ rbd status migration_target
2. Watchers: none
3. Migration:
4.     source: rbd/migration_source (5e2cba2f62e)
5.     destination: rbd/migration_target (5e2ed95ed806)
6.     state: prepared
```

Note that the source image will be moved to the RBD trash to avoid mistaken usage during the migration process:

```
1. $ rbd info migration_source
2. rbd: error opening image migration_source: (2) No such file or directory
3. $ rbd trash ls --all
4. 5e2cba2f62e migration_source
```

## Execute Migration

After preparing the live-migration, the image blocks from the source image must be copied to the target image. This is accomplished by running the rbd migration execute command:

```
1. $ rbd migration execute migration_target
2. Image migration: 100% complete...done.
```

The rbd status command will also provide feedback on the progress of the migration block deep-copy process:

```
1. $ rbd status migration_target
```

```

2. Watchers:
3.     watcher=1.2.3.4:0/3695551461 client.123 cookie=123
4. Migration:
5.     source: rbd/migration_source (5e2cba2f62e)
6.     destination: rbd/migration_target (5e2ed95ed806)
7.     state: executing (32% complete)

```

## Commit Migration

Once the live-migration has completed deep-copying all data blocks from the source image to the target, the migration can be committed:

```

1. $ rbd status migration_target
2. Watchers: none
3. Migration:
4.     source: rbd/migration_source (5e2cba2f62e)
5.     destination: rbd/migration_target (5e2ed95ed806)
6.     state: executed
7. $ rbd migration commit migration_target
8. Commit image migration: 100% complete...done.

```

If the migration\_source image is a parent of one or more clones, the -force option will need to be specified after ensuring all descendent clone images are not in use.

Committing the live-migration will remove the cross-links between the source and target images, and will remove the source image:

```
1. $ rbd trash list --all
```

## Abort Migration

If you wish to revert the prepare or execute step, run the rbd migration abort command to revert the migration process:

```

1. $ rbd migration abort migration_target
2. Abort image migration: 100% complete...done.

```

Aborting the migration will result in the target image being deleted and access to the original source image being restored:

```

1. $ rbd ls
2. migration_source

```

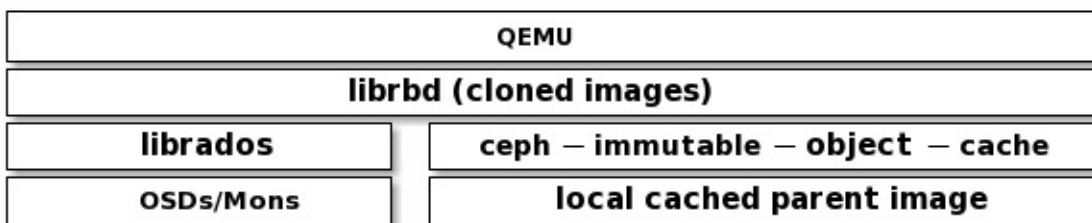
# RBD Persistent Cache

## Shared, Read-only Parent Image Cache

`Cloned RBD images` usually modify only a small fraction of the parent image. For example, in a VDI use-case, VMs are cloned from the same base image and initially differ only by hostname and IP address. During booting, all of these VMs read portions of the same parent image data. If we have a local cache of the parent image, this speeds up reads on the caching host. We also achieve reduction of client-to-cluster network traffic. RBD cache must be explicitly enabled in `ceph.conf`. The `ceph-immutable-object-cache` daemon is responsible for caching the parent content on the local disk, and future reads on that data will be serviced from the local cache.

### Note

RBD shared read-only parent image cache requires the Ceph Nautilus release or later.



## Enable RBD Shared Read-only Parent Image Cache

To enable RBD shared read-only parent image cache, the following Ceph settings need to be added in the `[client]` section of your `ceph.conf` file:

1. `rbd parent cache enabled = true`
2. `rbd plugins = parent_cache`

## Immutable Object Cache Daemon

The `ceph-immutable-object-cache` daemon is responsible for caching parent image content within its local caching directory. For better performance it's recommended to use SSDs as the underlying storage.

The key components of the daemon are:

1. **Domain socket based IPC:** The daemon will listen on a local domain socket on start

up and wait for connections from librbd clients.

2. **LRU based promotion/demotion policy:** The daemon will maintain in-memory statistics of cache-hits on each cache file. It will demote the cold cache if capacity reaches to the configured threshold.
3. **File-based caching store:** The daemon will maintain a simple file based cache store. On promotion the RADOS objects will be fetched from RADOS cluster and stored in the local caching directory.

On opening each cloned rbd image, `librbd` will try to connect to the cache daemon through its Unix domain socket. Once successfully connected, `librbd` will coordinate with the daemon on the subsequent reads. If there's a read that's not cached, the daemon will promote the RADOS object to local caching directory, so the next read on that object will be serviced from cache. The daemon also maintains simple LRU statistics so that under capacity pressure it will evict cold cache files as needed.

Here are some important cache configuration settings:

- `immutable_object_cache_sock` The path to the domain socket used for communication between librbd clients and the ceph-immutable-object-cache daemon.
- `immutable_object_cache_path` The immutable object cache data directory.
- `immutable_object_cache_max_size` The max size for immutable cache.
- `immutable_object_cache_watermark` The high-water mark for the cache. If the capacity reaches this threshold the daemon will delete cold cache based on LRU statistics.

The `ceph-immutable-object-cache` daemon is available within the optional `ceph-immutable-object-cache` distribution package.

### Important

`ceph-immutable-object-cache` daemon requires the ability to connect RADOS clusters.

`ceph-immutable-object-cache` daemon should use a unique Ceph user ID. To [create a Ceph user](#), with `ceph` specify the `auth get-or-create` command, user name, monitor caps, and OSD caps:

```
ceph auth get-or-create client.ceph-immutable-object-cache.{unique id} mon 'allow r' osd 'profile rbd-read-only'
```

The `ceph-immutable-object-cache` daemon can be managed by `systemd` by specifying the user ID as the daemon instance:

```
1. systemctl enable ceph-immutable-object-cache@immutable-object-cache.{unique id}
```

The `ceph-immutable-object-cache` can also be run in foreground by `ceph-immutable-object-cache` command:

```
1. ceph-immutable-object-cache -f --log-file={log_path}
```

# Config Settings

See [Block Device](#) for additional details.

## Generic IO Settings

`rbd compression hint`

Description

Hint to send to the OSDs on write operations. If set to compressible and the OSD bluestore compression mode setting is passive, the OSD will attempt to compress the data. If set to incompressible and the OSD compression setting is aggressive, the OSD will not attempt to compress the data.

Type

Enum

Required

No

Default

`none`

Values

`none` , `compressible` , `incompressible`

`rbd read from replica policy`

Description

policy for determining which OSD will receive read operations. If set to default, the primary OSD will always be used for read operations. If set to balance, read operations will be sent to a randomly selected OSD within the replica set. If set to localize, read operations will be sent to the closest OSD as determined by the CRUSH map. Note: this feature requires the cluster to be configured with a minimum compatible OSD release of Octopus.

Type

Enum

Required

No

## Default

default

## Values

default , balance , localize

# Cache Settings

## Kernel Caching

The kernel driver for Ceph block devices can use the Linux page cache to improve performance.

The user space implementation of the Ceph block device (i.e., librbd) cannot take advantage of the Linux page cache, so it includes its own in-memory caching, called “RBD caching.” RBD caching behaves just like well-behaved hard disk caching. When the OS sends a barrier or a flush request, all dirty data is written to the OSDs. This means that using write-back caching is just as safe as using a well-behaved physical hard disk with a VM that properly sends flushes (i.e. Linux kernel >= 2.6.32). The cache uses a Least Recently Used (LRU) algorithm, and in write-back mode it can coalesce contiguous requests for better throughput.

The librbd cache is enabled by default and supports three different cache policies: write-around, write-back, and write-through. Writes return immediately under both the write-around and write-back policies, unless there are more than rbd cache max dirty unwritten bytes to the storage cluster. The write-around policy differs from the write-back policy in that it does not attempt to service read requests from the cache, unlike the write-back policy, and is therefore faster for high performance write workloads. Under the write-through policy, writes return only when the data is on disk on all replicas, but reads may come from the cache.

Prior to receiving a flush request, the cache behaves like a write-through cache to ensure safe operation for older operating systems that do not send flushes to ensure crash consistent behavior.

If the librbd cache is disabled, writes and reads go directly to the storage cluster, and writes return only when the data is on disk on all replicas.

## Note

The cache is in memory on the client, and each RBD image has its own. Since the cache is local to the client, there’s no coherency if there are others accessing the image. Running GFS or OCFS on top of RBD will not work with caching enabled.

The ceph.conf file settings for RBD should be set in the [client] section of your configuration file. The settings include:

rbd cache

## Description

Enable caching for RADOS Block Device (RBD).

## Type

Boolean

## Required

No

## Default

`true`

`rbd cache policy`

## Description

Select the caching policy for librbd.

## Type

Enum

## Required

No

## Default

`writearound`

## Values

`writearound` , `writeback` , `writethrough`

`rbd cache writethrough until flush`

## Description

Start out in write-through mode, and switch to write-back after the first flush request is received. Enabling this is a conservative but safe setting in case VMs running on rbd are too old to send flushes, like the virtio driver in Linux before 2.6.32.

## Type

Boolean

## Required

No

## Default

```
true  
rbd cache size
```

## Description

The RBD cache size in bytes.

## Type

64-bit Integer

## Required

No

## Default

```
32 MiB
```

## Policies

write-back and write-through

```
rbd cache max dirty
```

## Description

The `dirty` limit in bytes at which the cache triggers write-back. If `0`, uses write-through caching.

## Type

64-bit Integer

## Required

No

## Constraint

Must be less than `rbd cache size`.

## Default

```
24 MiB
```

## Policies

write-around and write-back

```
rbd cache target dirty
```

## Description

The `dirty target` before the cache begins writing data to the data storage. Does not block writes to the cache.

Type

64-bit Integer

Required

No

Constraint

Must be less than `rbd cache max dirty`.

Default

`16 MiB`

Policies

write-back

`rbd cache max dirty age`

Description

The number of seconds dirty data is in the cache before writeback starts.

Type

Float

Required

No

Default

`1.0`

Policies

write-back

## Read-ahead Settings

librbd supports read-ahead/prefetching to optimize small, sequential reads. This should normally be handled by the guest OS in the case of a VM, but boot loaders may not issue efficient reads. Read-ahead is automatically disabled if caching is disabled or if the policy is write-around.

`rbd readahead trigger requests`

## Description

Number of sequential read requests necessary to trigger read-ahead.

## Type

Integer

## Required

No

## Default

10

```
rbd readahead max bytes
```

## Description

Maximum size of a read-ahead request. If zero, read-ahead is disabled.

## Type

64-bit Integer

## Required

No

## Default

512 KiB

```
rbd readahead disable after bytes
```

## Description

After this many bytes have been read from an RBD image, read-ahead is disabled for that image until it is closed. This allows the guest OS to take over read-ahead once it is booted. If zero, read-ahead stays enabled.

## Type

64-bit Integer

## Required

No

## Default

50 MiB

# Image Features

RBD supports advanced features which can be specified via the command line when creating images or the default features can be specified via Ceph config file via ‘rbd\_default\_features = <sum of feature numeric values>’ or ‘rbd\_default\_features = <comma-delimited list of CLI values>’

#### Layering

##### Description

Layering enables you to use cloning.

##### Internal value

1

##### CLI value

layering

##### Added in

v0.52 (Bobtail)

##### KRBD support

since v3.10

##### Default

yes

#### Striping v2

##### Description

Striping spreads data across multiple objects. Striping helps with parallelism for sequential read/write workloads.

##### Internal value

2

##### CLI value

striping

##### Added in

v0.55 (Bobtail)

##### KRBD support

since v3.10 (default striping only, “fancy” striping added in v4.17)

Default

yes

`Exclusive locking`

Description

When enabled, it requires a client to get a lock on an object before making a write. Exclusive lock should only be enabled when a single client is accessing an image at the same time.

Internal value

4

CLI value

`exclusive-lock`

Added in

v0.92 (Hammer)

KRBD support

since v4.9

Default

yes

`Object map`

Description

Object map support depends on exclusive lock support. Block devices are thin provisioned—meaning, they only store data that actually exists. Object map support helps track which objects actually exist (have data stored on a drive). Enabling object map support speeds up I/O operations for cloning; importing and exporting a sparsely populated image; and deleting.

Internal value

8

CLI value

`object-map`

Added in

v0.93 (Hammer)

KRBD support

since v5.3

Default

yes

Fast-diff

Description

Fast-diff support depends on object map support and exclusive lock support. It adds another property to the object map, which makes it much faster to generate diffs between snapshots of an image, and the actual data usage of a snapshot much faster.

Internal value

16

CLI value

fast-diff

Added in

v9.0.1 (Infernalis)

KRBD support

since v5.3

Default

yes

Deep-flatten

Description

Deep-flatten makes rbd flatten work on all the snapshots of an image, in addition to the image itself. Without it, snapshots of an image will still rely on the parent, so the parent will not be delete-able until the snapshots are deleted. Deep-flatten makes a parent independent of its clones, even if they have snapshots.

Internal value

32

CLI value

deep-flatten

Added in

## v9.0.2 (Infernalis)

KRBD support

since v5.1

Default

yes

**Journaling**

### Description

Journaling support depends on exclusive lock support. Journaling records all modifications to an image in the order they occur. RBD mirroring utilizes the journal to replicate a crash consistent image to a remote cluster.

Internal value

64

CLI value

journaling

Added in

v10.0.1 (Jewel)

KRBD support

no

Default

no

**Data pool**

### Description

On erasure-coded pools, the image data block objects need to be stored on a separate pool from the image metadata.

Internal value

128

Added in

v11.1.0 (Kraken)

KRBD support

since v4.11

Default

no

Operations

Description

Used to restrict older clients from performing certain maintenance operations against an image (e.g. clone, snap create).

Internal value

256

Added in

v13.0.2 (Mimic)

KRBD support

since v4.16

Migrating

Description

Used to restrict older clients from opening an image when it is in migration state.

Internal value

512

Added in

v14.0.1 (Nautilus)

KRBD support

no

Non-primary

Description

Used to restrict changes to non-primary images using snapshot-based mirroring.

Internal value

1024

Added in

v15.2.0 (Octopus)

KRBD support

no

## QOS Settings

---

librbd supports limiting per image IO, controlled by the following settings.

`rbd qos iops limit`

Description

The desired limit of IO operations per second.

Type

Unsigned Integer

Required

No

Default

`0`

`rbd qos bps limit`

Description

The desired limit of IO bytes per second.

Type

Unsigned Integer

Required

No

Default

`0`

`rbd qos read iops limit`

Description

The desired limit of read operations per second.

Type

**Unsigned Integer**

Required

No

**Default**

0

```
rbd qos write iops limit
```

**Description**

The desired limit of write operations per second.

**Type**

**Unsigned Integer**

Required

No

**Default**

0

```
rbd qos read bps limit
```

**Description**

The desired limit of read bytes per second.

**Type**

**Unsigned Integer**

Required

No

**Default**

0

```
rbd qos write bps limit
```

**Description**

The desired limit of write bytes per second.

**Type**

**Unsigned Integer**

**Required**

No

**Default**

0

rbd qos iops burst

**Description**

The desired burst limit of IO operations.

**Type**

Unsigned Integer

**Required**

No

**Default**

0

rbd qos bps burst

**Description**

The desired burst limit of IO bytes.

**Type**

Unsigned Integer

**Required**

No

**Default**

0

rbd qos read iops burst

**Description**

The desired burst limit of read operations.

**Type**

Unsigned Integer

**Required**

No

Default

0

rbd qos write iops burst

Description

The desired burst limit of write operations.

Type

Unsigned Integer

Required

No

Default

0

rbd qos read bps burst

Description

The desired burst limit of read bytes.

Type

Unsigned Integer

Required

No

Default

0

rbd qos write bps burst

Description

The desired burst limit of write bytes.

Type

Unsigned Integer

Required

No

**Default**

0

```
rbd qos iops burst seconds
```

**Description**

The desired burst duration in seconds of IO operations.

**Type**

Unsigned Integer

**Required**

No

**Default**

1

```
rbd qos bps burst seconds
```

**Description**

The desired burst duration in seconds of IO bytes.

**Type**

Unsigned Integer

**Required**

No

**Default**

1

```
rbd qos read iops burst seconds
```

**Description**

The desired burst duration in seconds of read operations.

**Type**

Unsigned Integer

**Required**

No

**Default**

`1``rbd qos write iops burst seconds`**Description**

The desired burst duration in seconds of write operations.

**Type**

Unsigned Integer

**Required**

No

**Default**`1``rbd qos read bps burst seconds`**Description**

The desired burst duration in seconds of read bytes.

**Type**

Unsigned Integer

**Required**

No

**Default**`1``rbd qos write bps burst seconds`**Description**

The desired burst duration in seconds of write bytes.

**Type**

Unsigned Integer

**Required**

No

**Default**`1``rbd qos schedule tick min`

## Description

The minimum schedule tick (in milliseconds) for QoS.

## Type

Unsigned Integer

## Required

No

## Default

50

# RBD Replay

RBD Replay is a set of tools for capturing and replaying RADOS Block Device (RBD) workloads. To capture an RBD workload, `lttng-tools` must be installed on the client, and `librbd` on the client must be the v0.87 (Giant) release or later. To replay an RBD workload, `librbd` on the client must be the Giant release or later.

Capture and replay takes three steps:

1. Capture the trace. Make sure to capture `pthread_id` context:

```
1. mkdir -p traces
2. lttng create -o traces librbd
3. lttng enable-event -u 'librbd:/*'
4. lttng add-context -u -t pthread_id
5. lttng start
6. # run RBD workload here
7. lttng stop
```

2. Process the trace with `rbd-replay-prep`:

```
1. rbd-replay-prep traces/ust/uid/*/* replay.bin
```

3. Replay the trace with `rbd-replay`. Use read-only until you know it's doing what you want:

```
1. rbd-replay --read-only replay.bin
```

## Important

`rbd-replay` will destroy data by default. Do not use against an image you wish to keep, unless you use the `--read-only` option.

The replayed workload does not have to be against the same RBD image or even the same cluster as the captured workload. To account for differences, you may need to use the `--pool` and `--map-image` options of `rbd-replay`.

# Kernel Module Operations

## Important

To use kernel module operations, you must have a running Ceph cluster.

## Get a List of Images

To mount a block device image, first return a list of the images.

```
1. rbd list
```

## Map a Block Device

Use `rbd` to map an image name to a kernel module. You must specify the image name, the pool name, and the user name. `rbd` will load RBD kernel module on your behalf if it's not already loaded.

```
1. sudo rbd device map {pool-name}/{image-name} --id {user-name}
```

For example:

```
1. sudo rbd device map rbd/myimage --id admin
```

If you use `cephx` authentication, you must also specify a secret. It may come from a keyring or a file containing the secret.

```
1. sudo rbd device map rbd/myimage --id admin --keyring /path/to/keyring
2. sudo rbd device map rbd/myimage --id admin --keyfile /path/to/file
```

## Show Mapped Block Devices

To show block device images mapped to kernel modules with the `rbd`, specify `device list` arguments.

```
1. rbd device list
```

## Unmapping a Block Device

To unmap a block device image with the `rbd` command, specify the `device unmap` arguments and the device name (i.e., by convention the same as the block device image

name).

```
1. sudo rbd device unmap /dev/rbd/{poolname}/{imagename}
```

For example:

```
1. sudo rbd device unmap /dev/rbd/rbd/foo
```

# QEMU and Block Devices

The most frequent Ceph Block Device use case involves providing block device images to virtual machines. For example, a user may create a “golden” image with an OS and any relevant software in an ideal configuration. Then the user takes a snapshot of the image. Finally the user clones the snapshot (potentially many times). See [Snapshots](#) for details. The ability to make copy-on-write clones of a snapshot means that Ceph can provision block device images to virtual machines quickly, because the client doesn’t have to download the entire image each time it spins up a new virtual machine.



Ceph Block Devices attach to QEMU virtual machines. For details on QEMU, see [QEMU Open Source Processor Emulator](#). For QEMU documentation, see [QEMU Manual](#). For installation details, see [Installation](#).

## Important

To use Ceph Block Devices with QEMU, you must have access to a running Ceph cluster.

## Usage

The QEMU command line expects you to specify the Ceph pool and image name. You may also specify a snapshot.

QEMU will assume that Ceph configuration resides in the default location (e.g., `/etc/ceph/$cluster.conf`) and that you are executing commands as the default `client.admin` user unless you expressly specify another Ceph configuration file path or another user. When specifying a user, QEMU uses the `ID` rather than the full `TYPE:ID`. See [User Management - User](#) for details. Do not prepend the client type (i.e., `client.`) to the beginning of the user `ID`, or you will receive an authentication error. You should have the key for the `admin` user or the key of another user you specify with the `:id={user}` option in a keyring file stored in default path (i.e., `/etc/ceph` or the local directory with appropriate file ownership and permissions. Usage takes the following form:

```
1. qemu-img {command} [options] rbd:{pool-name}/{image-name}[@snapshot-name][:option1=value1][:option2=value2...]
```

For example, specifying the `id` and `conf` options might look like the following:

```
1. qemu-img {command} [options] rbd:glance-pool/maipo:id=glance:conf=/etc/ceph/ceph.conf
```

### Tip

Configuration values containing `:`, `@`, or `=` can be escaped with a leading `\` character.

## Creating Images with QEMU

You can create a block device image from QEMU. You must specify `rbd`, the pool name, and the name of the image you wish to create. You must also specify the size of the image.

```
1. qemu-img create -f raw rbd:{pool-name}/{image-name} {size}
```

For example:

```
1. qemu-img create -f raw rbd:data/foo 10G
```

### Important

The `raw` data format is really the only sensible `format` option to use with RBD. Technically, you could use other QEMU-supported formats (such as `qcow2` or `vmdk`), but doing so would add additional overhead, and would also render the volume unsafe for virtual machine live migration when caching (see below) is enabled.

## Resizing Images with QEMU

You can resize a block device image from QEMU. You must specify `rbd`, the pool name, and the name of the image you wish to resize. You must also specify the size of the image.

```
1. qemu-img resize rbd:{pool-name}/{image-name} {size}
```

For example:

```
1. qemu-img resize rbd:data/foo 10G
```

## Retrieving Image Info with QEMU

You can retrieve block device image information from QEMU. You must specify `rbd`, the pool name, and the name of the image.

```
1. qemu-img info rbd:{pool-name}/{image-name}
```

For example:

```
1. qemu-img info rbd:data/foo
```

## Running QEMU with RBD

QEMU can pass a block device from the host on to a guest, but since QEMU 0.15, there's no need to map an image as a block device on the host. Instead, QEMU attaches an image as a virtual block device directly via `librbd`. This strategy increases performance by avoiding context switches and taking advantage of [RBD caching](#).

You can use `qemu-img` to convert existing virtual machine images to Ceph block device images. For example, if you have a qcow2 image, you could run:

```
1. qemu-img convert -f qcow2 -O raw debian_squeeze.qcow2 rbd:data/squeeze
```

To run a virtual machine booting from that image, you could run:

```
1. qemu -m 1024 -drive format=raw,file=rbd:data/squeeze
```

[RBD caching](#) can significantly improve performance. Since QEMU 1.2, QEMU's cache options control `librbd` caching:

```
1. qemu -m 1024 -drive format=rbd,file=rbd:data/squeeze,cache=writeback
```

If you have an older version of QEMU, you can set the `librbd` cache configuration (like any Ceph configuration option) as part of the 'file' parameter:

```
1. qemu -m 1024 -drive format=raw,file=rbd:data/squeeze:rbd_cache=true,cache=writeback
```

### Important

If you set `rbd_cache=true`, you must set `cache=writeback` or risk data loss. Without `cache=writeback`, QEMU will not send flush requests to librbd. If QEMU exits uncleanly in this configuration, file systems on top of rbd can be corrupted.

## Enabling Discard/TRIM

Since Ceph version 0.46 and QEMU version 1.1, Ceph Block Devices support the discard operation. This means that a guest can send TRIM requests to let a Ceph block device reclaim unused space. This can be enabled in the guest by mounting `ext4` or `XFS` with the `discard` option.

For this to be available to the guest, it must be explicitly enabled for the block device. To do this, you must specify a `discard_granularity` associated with the drive:

```
1. qemu -m 1024 -drive format=raw,file=rbd:data/squeeze,id=drive1,if=none \
2.      -device ide-hd,drive=drive1,discard_granularity=512
```

Note that this uses the IDE driver. The virtio driver does not support discard.

If using libvirt, edit your libvirt domain's configuration file using `virsh edit` to include the `xmlns:qemu` value. Then, add a `<qemu:commandline>` block as a child of that domain. The following example shows how to set two devices with `qemu id=` to different `discard_granularity` values.

```
1. <domain type='kvm' xmlns:qemu='http://libvirt.org/schemas/domain/qemu/1.0'>
2.   <qemu:commandline>
3.     <qemu:arg value=' -set' />
4.     <qemu:arg value='block.scsi0-0-0.discard_granularity=4096' />
5.     <qemu:arg value=' -set' />
6.     <qemu:arg value='block.scsi0-0-1.discard_granularity=65536' />
7.   </qemu:commandline>
8. </domain>
```

## QEMU Cache Options

QEMU's cache options correspond to the following Ceph [RBD Cache](#) settings.

Writeback:

```
1. rbd_cache = true
```

Writethrough:

```
1. rbd_cache = true
2. rbd_cache_max_dirty = 0
```

None:

```
1. rbd_cache = false
```

QEMU's cache settings override Ceph's cache settings (including settings that are explicitly set in the Ceph configuration file).

Note

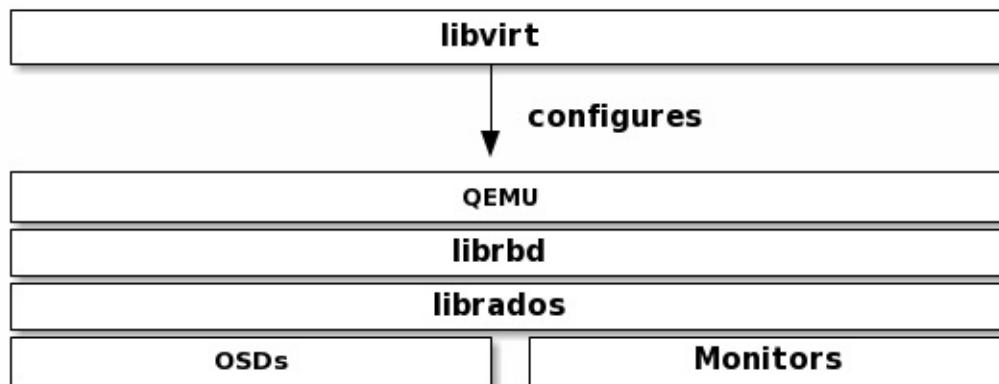
Prior to QEMU v2.4.0, if you explicitly set [RBD Cache](#) settings in the Ceph configuration file, your Ceph settings override the QEMU cache settings.

# Using libvirt with Ceph RBD

The `libvirt` library creates a virtual machine abstraction layer between hypervisor interfaces and the software applications that use them. With `libvirt`, developers and system administrators can focus on a common management framework, common API, and common shell interface (i.e., `virsh`) to many different hypervisors, including:

- QEMU/KVM
- XEN
- LXC
- VirtualBox
- etc.

Ceph block devices support QEMU/KVM. You can use Ceph block devices with software that interfaces with `libvirt`. The following stack diagram illustrates how `libvirt` and QEMU use Ceph block devices via `librbd`.



The most common `libvirt` use case involves providing Ceph block devices to cloud solutions like OpenStack or CloudStack. The cloud solution uses `libvirt` to interact with QEMU/KVM, and QEMU/KVM interacts with Ceph block devices via `librbd`. See [Block Devices and OpenStack](#) and [Block Devices and CloudStack](#) for details. See [Installation](#) for installation details.

You can also use Ceph block devices with `libvirt`, `virsh` and the `libvirt` API. See [libvirt Virtualization API](#) for details.

To create VMs that use Ceph block devices, use the procedures in the following sections. In the exemplary embodiment, we have used `libvirt-pool` for the pool name, `client.libvirt` for the user name, and `new-libvirt-image` for the image name. You may use any value you like, but ensure you replace those values when executing commands in the

subsequent procedures.

## Configuring Ceph

To configure Ceph for use with `libvirt`, perform the following steps:

1. [Create a pool](#). The following example uses the pool name `libvirt-pool` .:

```
1. ceph osd pool create libvirt-pool
```

Verify the pool exists.

```
1. ceph osd lspools
```

2. Use the `rbd` tool to initialize the pool for use by RBD:

```
1. rbd pool init <pool-name>
```

3. [Create a Ceph User](#) (or use `client.admin` for version 0.9.7 and earlier). The following example uses the Ceph user name `client.libvirt` and references `libvirt-pool` .

```
1. ceph auth get-or-create client.libvirt mon 'profile rbd' osd 'profile rbd pool=libvirt-pool'
```

Verify the name exists.

```
1. ceph auth ls
```

**NOTE:** `libvirt` will access Ceph using the ID `libvirt`, not the Ceph name `client.libvirt`. See [User Management - User](#) and [User Management - CLI](#) for a detailed explanation of the difference between ID and name.

4. Use QEMU to [create an image](#) in your RBD pool. The following example uses the image name `new-libvirt-image` and references `libvirt-pool` .

```
1. qemu-img create -f rbd rbd:libvirt-pool/new-libvirt-image 2G
```

Verify the image exists.

```
1. rbd -p libvirt-pool ls
```

**NOTE:** You can also use `rbd create` to create an image, but we recommend ensuring that QEMU is working properly.

Tip

Optionally, if you wish to enable debug logs and the admin socket for this client, you can add the following section to `/etc/ceph/ceph.conf` :

```
1. [client.libvirt]
2. log file = /var/log/ceph/qemu-guest-$pid.log
3. admin socket = /var/run/ceph/$cluster-$type.$id.$pid.$cctid.asok
```

The `client.libvirt` section name should match the cephx user you created above. If SELinux or AppArmor is enabled, note that this could prevent the client process (qemu via libvirt) from doing some operations, such as writing logs or operate the images or admin socket to the destination locations (`/var/log/ceph` or `/var/run/ceph`). Additionally, make sure that the libvirt and qemu users have appropriate access to the specified directory.

## Preparing the VM Manager

You may use `libvirt` without a VM manager, but you may find it simpler to create your first domain with `virt-manager`.

1. Install a virtual machine manager. See [KVM/VirtManager](#) for details.

```
1. sudo apt-get install virt-manager
```

2. Download an OS image (if necessary).
3. Launch the virtual machine manager.

```
1. sudo virt-manager
```

## Creating a VM

To create a VM with `virt-manager`, perform the following steps:

1. Press the **Create New Virtual Machine** button.
2. Name the new virtual machine domain. In the exemplary embodiment, we use the name `libvirt-virtual-machine`. You may use any name you wish, but ensure you replace `libvirt-virtual-machine` with the name you choose in subsequent commandline and configuration examples.

```
1. libvirt-virtual-machine
```

3. Import the image.

```
1. /path/to/image/recent-linux.img
```

**NOTE:** Import a recent image. Some older images may not rescan for virtual devices properly.

4. Configure and start the VM.
5. You may use `virsh list` to verify the VM domain exists.

```
1. sudo virsh list
```

6. Login to the VM (root/root)
7. Stop the VM before configuring it for use with Ceph.

## Configuring the VM

When configuring the VM for use with Ceph, it is important to use `virsh` where appropriate. Additionally, `virsh` commands often require root privileges (i.e., `sudo`) and will not return appropriate results or notify you that root privileges are required. For a reference of `virsh` commands, refer to [Virsh Command Reference](#).

1. Open the configuration file with `virsh edit`.

```
1. sudo virsh edit {vm-domain-name}
```

Under `<devices>` there should be a `<disk>` entry.

```
1. <devices>
2.   <emulator>/usr/bin/kvm</emulator>
3.   <disk type='file' device='disk'>
4.     <driver name='qemu' type='raw' />
5.     <source file='/path/to/image/recent-linux.img' />
6.     <target dev='vda' bus='virtio' />
7.     <address type='drive' controller='0' bus='0' unit='0' />
8.   </disk>
```

Replace `/path/to/image/recent-linux.img` with the path to the OS image. The minimum kernel for using the faster `virtio` bus is 2.6.25. See [Virtio](#) for details.

**IMPORTANT:** Use `sudo virsh edit` instead of a text editor. If you edit the configuration file under `/etc/libvirt/qemu` with a text editor, `libvirt` may not recognize the change. If there is a discrepancy between the contents of the XML file under `/etc/libvirt/qemu` and the result of `sudo virsh dumpxml {vm-domain-name}`, then your VM may not work properly.

2. Add the Ceph RBD image you created as a `<disk>` entry.

```
1. <disk type='network' device='disk'>
2.   <source protocol='rbd' name='libvirt-pool/new-libvirt-image' />
```

```

3.           <host name='{monitor-host}' port='6789' />
4.       </source>
5.       <target dev='vdb' bus='virtio' />
6.   </disk>

```

Replace `{monitor-host}` with the name of your host, and replace the pool and/or image name as necessary. You may add multiple `<host>` entries for your Ceph monitors. The `dev` attribute is the logical device name that will appear under the `/dev` directory of your VM. The optional `bus` attribute indicates the type of disk device to emulate. The valid settings are driver specific (e.g., “ide”, “scsi”, “virtio”, “xen”, “usb” or “sata”).

See [Disks](#) for details of the `<disk>` element, and its child elements and attributes.

3. Save the file.
4. If your Ceph Storage Cluster has [Ceph Authentication](#) enabled (it does by default), you must generate a secret.

```

1. cat > secret.xml <<EOF
2. <secret ephemeral='no' private='no'>
3.   <usage type='ceph'>
4.     <name>client.libvirt secret</name>
5.   </usage>
6. </secret>
7. EOF

```

5. Define the secret.

```

1. sudo virsh secret-define --file secret.xml
2. {uuid of secret}

```

6. Get the `client.libvirt` key and save the key string to a file.

```
1. ceph auth get-key client.libvirt | sudo tee client.libvirt.key
```

7. Set the UUID of the secret.

```

sudo virsh secret-set-value --secret {uuid of secret} --base64 $(cat client.libvirt.key) && rm
1. client.libvirt.key secret.xml

```

You must also set the secret manually by adding the following `<auth>` entry to the `<disk>` element you entered earlier (replacing the `uuid` value with the result from the command line example above).

```
1. sudo virsh edit {vm-domain-name}
```

Then, add `<auth></auth>` element to the domain configuration file:

```

1. ...
2. </source>
3. <auth username='libvirt'>
4.     <secret type='ceph' uuid='{uuid of secret}'/>
5. </auth>
6. <target ...>

```

**NOTE:** The exemplary ID is `libvirt`, not the Ceph name `client.libvirt` as generated at step 2 of [Configuring Ceph](#). Ensure you use the ID component of the Ceph name you generated. If for some reason you need to regenerate the secret, you will have to execute `sudo virsh secret-undefine {uuid}` before executing `sudo virsh secret-set-value` again.

## Summary

Once you have configured the VM for use with Ceph, you can start the VM. To verify that the VM and Ceph are communicating, you may perform the following procedures.

1. Check to see if Ceph is running:

```
1. ceph health
```

2. Check to see if the VM is running.

```
1. sudo virsh list
```

3. Check to see if the VM is communicating with Ceph. Replace `{vm-domain-name}` with the name of your VM domain:

```
1. sudo virsh qemu-monitor-command --hmp {vm-domain-name} 'info block'
```

4. Check to see if the device from `<target dev='vdb' bus='virtio' />` exists:

```
1. virsh domblklist {vm-domain-name} --details
```

If everything looks okay, you may begin using the Ceph block device within your VM.

# Block Devices and Kubernetes

You may use Ceph Block Device images with Kubernetes v1.13 and later through [ceph-csi](#), which dynamically provisions RBD images to back Kubernetes [volumes](#) and maps these RBD images as block devices (optionally mounting a file system contained within the image) on worker nodes running [pods](#) that reference an RBD-backed volume. Ceph stripes block device images as objects across the cluster, which means that large Ceph Block Device images have better performance than a standalone server!

To use Ceph Block Devices with Kubernetes v1.13 and higher, you must install and configure [ceph-csi](#) within your Kubernetes environment. The following diagram depicts the Kubernetes/Ceph technology stack.

## Important

[ceph-csi](#) uses the RBD kernel modules by default which may not support all Ceph [CRUSH tunables](#) or [RBD image features](#).

## Create a Pool

By default, Ceph block devices use the [rbd](#) pool. Create a pool for Kubernetes volume storage. Ensure your Ceph cluster is running, then create the pool.

```
1. $ ceph osd pool create kubernetes
```

See [Create a Pool](#) for details on specifying the number of placement groups for your pools, and [Placement Groups](#) for details on the number of placement groups you should set for your pools.

A newly created pool must be initialized prior to use. Use the [rbd](#) tool to initialize the pool:

```
1. $ rbd pool init kubernetes
```

## Configure ceph-csi

## Setup Ceph Client Authentication

Create a new user for Kubernetes and ceph-csi. Execute the following and record the generated key:

```
$ ceph auth get-or-create client.kubernetes mon 'profile rbd' osd 'profile rbd pool=kubernetes' mgr 'profile rbd pool=kubernetes'  
1. [client.kubernetes]  
2. [client.kubernetes]
```

```
3.     key = AQD9o0Fd6hQRChAA7fMaSZXduT3NWEqy1Npmg==
```

## Generate ceph-csi ConfigMap

The ceph-csi requires a ConfigMap object stored in Kubernetes to define the the Ceph monitor addresses for the Ceph cluster. Collect both the Ceph cluster unique fsid and the monitor addresses:

```
1. $ ceph mon dump
2. <...>
3. fsid b9127830-b0cc-4e34-aa47-9d1a2e9949a8
4. <...>
5. 0: [v2:192.168.1.1:3300/0, v1:192.168.1.1:6789/0] mon.a
6. 1: [v2:192.168.1.2:3300/0, v1:192.168.1.2:6789/0] mon.b
7. 2: [v2:192.168.1.3:3300/0, v1:192.168.1.3:6789/0] mon.c
```

### Note

`ceph-csi` currently only supports the [legacy V1 protocol](#).

Generate a `csi-config-map.yaml` file similar to the example below, substituting the `fsid` for “`clusterID`”, and the monitor addresses for “`monitors`”:

```
1. $ cat <<EOF > csi-config-map.yaml
2. ---
3. apiVersion: v1
4. kind: ConfigMap
5. data:
6.   config.json: |-
7.     [
8.       {
9.         "clusterID": "b9127830-b0cc-4e34-aa47-9d1a2e9949a8",
10.        "monitors": [
11.          "192.168.1.1:6789",
12.          "192.168.1.2:6789",
13.          "192.168.1.3:6789"
14.        ]
15.      }
16.    ]
17. metadata:
18.   name: ceph-csi-config
19. EOF
```

Once generated, store the new ConfigMap object in Kubernetes:

```
1. $ kubectl apply -f csi-config-map.yaml
```

## Generate ceph-csi cephx Secret

ceph-csi requires the cephx credentials for communicating with the Ceph cluster. Generate a `csi-rbd-secret.yaml` file similar to the example below, using the newly created Kubernetes user id and cephx key:

```

1. $ cat <<EOF > csi-rbd-secret.yaml
2. ---
3. apiVersion: v1
4. kind: Secret
5. metadata:
6.   name: csi-rbd-secret
7.   namespace: default
8.   stringData:
9.     userID: kubernetes
10.    userKey: AQD9o0Fd6hQRChAA7fMaSZXduT3NWEqylNpmg==
11. EOF

```

Once generated, store the new Secret object in Kubernetes:

```
1. $ kubectl apply -f csi-rbd-secret.yaml
```

## Configure ceph-csi Plugins

Create the required ServiceAccount and RBAC ClusterRole/ClusterRoleBinding Kubernetes objects. These objects do not necessarily need to be customized for your Kubernetes environment and therefore can be used as-is from the ceph-csi deployment YAMLS:

```

$ kubectl apply -f https://raw.githubusercontent.com/ceph/ceph-csi/master/deploy/rbd/kubernetes/csi-
1. provisioner-rbac.yaml
$ kubectl apply -f https://raw.githubusercontent.com/ceph/ceph-csi/master/deploy/rbd/kubernetes/csi-
2. nodeplugin-rbac.yaml

```

Finally, create the ceph-csi provisioner and node plugins. With the possible exception of the ceph-csi container release version, these objects do not necessarily need to be customized for your Kubernetes environment and therefore can be used as-is from the ceph-csi deployment YAMLS:

```

$ wget https://raw.githubusercontent.com/ceph/ceph-csi/master/deploy/rbd/kubernetes/csi-rbdplugin-
1. provisioner.yaml
2. $ kubectl apply -f csi-rbdplugin-provisioner.yaml
3. $ wget https://raw.githubusercontent.com/ceph/ceph-csi/master/deploy/rbd/kubernetes/csi-rbdplugin.yaml
4. $ kubectl apply -f csi-rbdplugin.yaml

```

### Important

The provisioner and node plugin YAMLS will, by default, pull the development release of the ceph-csi container (`quay.io/cephcsi/cephcsi:canary`). The YAMLS should be updated to use a release version container for production workloads.

# Using Ceph Block Devices

## Create a StorageClass

The Kubernetes StorageClass defines a class of storage. Multiple StorageClass objects can be created to map to different quality-of-service levels (i.e. NVMe vs HDD-based pools) and features.

For example, to create a ceph-csi StorageClass that maps to the kubernetes pool created above, the following YAML file can be used after ensuring that the “clusterID” property matches your Ceph cluster’s fsid:

```
1. $ cat <<EOF > csi-rbd-sc.yaml
2. ---
3. apiVersion: storage.k8s.io/v1
4. kind: StorageClass
5. metadata:
6.   name: csi-rbd-sc
7.   provisioner: rbd.csi.ceph.com
8.   parameters:
9.     clusterID: b9127830-b0cc-4e34-aa47-9d1a2e9949a8
10.    pool: kubernetes
11.    csi.storage.k8s.io/provisioner-secret-name: csi-rbd-secret
12.    csi.storage.k8s.io/provisioner-secret-namespace: default
13.    csi.storage.k8s.io/node-stage-secret-name: csi-rbd-secret
14.    csi.storage.k8s.io/node-stage-secret-namespace: default
15.   reclaimPolicy: Delete
16.   mountOptions:
17.     - discard
18. EOF
19. $ kubectl apply -f csi-rbd-sc.yaml
```

## Create a PersistentVolumeClaim

A PersistentVolumeClaim is a request for abstract storage resources by a user. The PersistentVolumeClaim would then be associated to a Pod resource to provision a PersistentVolume, which would be backed by a Ceph block image. An optional volumeMode can be included to select between a mounted file system (default) or raw block device-based volume.

Using ceph-csi, specifying Filesystem for volumeMode can support both ReadWriteOnce and ReadOnlyMany accessMode claims, and specifying Block for volumeMode can support ReadWriteOnce, ReadWriteMany, and ReadOnlyMany accessMode claims.

For example, to create a block-based PersistentVolumeClaim that utilizes the ceph-csi-based StorageClass created above, the following YAML can be used to request raw block storage from the csi-rbd-sc StorageClass:

```

1. $ cat <<EOF > raw-block-pvc.yaml
2. ---
3. apiVersion: v1
4. kind: PersistentVolumeClaim
5. metadata:
6.   name: raw-block-pvc
7. spec:
8.   accessModes:
9.     - ReadWriteOnce
10.  volumeMode: Block
11.  resources:
12.    requests:
13.      storage: 1Gi
14.  storageClassName: csi-rbd-sc
15. EOF
16. $ kubectl apply -f raw-block-pvc.yaml

```

The following demonstrates an example of binding the above PersistentVolumeClaim to a Pod resource as a raw block device:

```

1. $ cat <<EOF > raw-block-pod.yaml
2. ---
3. apiVersion: v1
4. kind: Pod
5. metadata:
6.   name: pod-with-raw-block-volume
7. spec:
8.   containers:
9.     - name: fc-container
10.       image: fedora:26
11.       command: ["/bin/sh", "-c"]
12.       args: ["tail -f /dev/null"]
13.       volumeDevices:
14.         - name: data
15.           devicePath: /dev/xvda
16.       volumes:
17.         - name: data
18.           persistentVolumeClaim:
19.             claimName: raw-block-pvc
20. EOF
21. $ kubectl apply -f raw-block-pod.yaml

```

To create a file-system-based PersistentVolumeClaim that utilizes the ceph-csi-based StorageClass created above, the following YAML can be used to request a mounted file system (backed by an RBD image) from the csi-rbd-sc StorageClass:

```

1. $ cat <<EOF > pvc.yaml
2. ---
3. apiVersion: v1
4. kind: PersistentVolumeClaim
5. metadata:

```

```
6.   name: rbd-pvc
7. spec:
8.   accessModes:
9.     - ReadWriteOnce
10.  volumeMode: Filesystem
11.  resources:
12.    requests:
13.      storage: 1Gi
14.  storageClassName: csi-rbd-sc
15. EOF
16. $ kubectl apply -f pvc.yaml
```

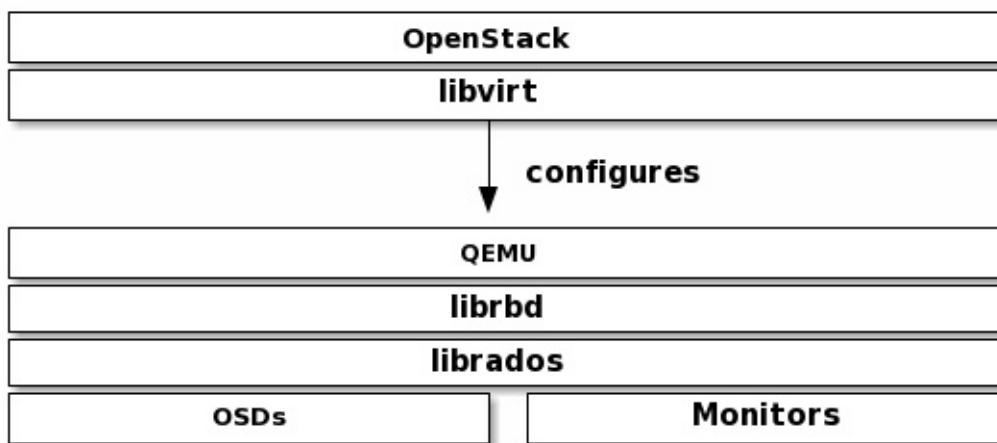
The following demonstrates and example of binding the above PersistentVolumeClaim to a Pod resource as a mounted file system:

```
1. $ cat <<EOF > pod.yaml
2. ---
3. apiVersion: v1
4. kind: Pod
5. metadata:
6.   name: csi-rbd-demo-pod
7. spec:
8.   containers:
9.     - name: web-server
10.    image: nginx
11.    volumeMounts:
12.      - name: mypvc
13.        mountPath: /var/lib/www/html
14.    volumes:
15.      - name: mypvc
16.        persistentVolumeClaim:
17.          claimName: rbd-pvc
18.          readOnly: false
19. EOF
20. $ kubectl apply -f pod.yaml
```

# Block Devices and OpenStack

You can attach Ceph Block Device images to OpenStack instances through `libvirt`, which configures the QEMU interface to `librbd`. Ceph stripes block volumes across multiple OSDs within the cluster, which means that large volumes can realize better performance than local drives on a standalone server!

To use Ceph Block Devices with OpenStack, you must install QEMU, `libvirt`, and OpenStack first. We recommend using a separate physical node for your OpenStack installation. OpenStack recommends a minimum of 8GB of RAM and a quad-core processor. The following diagram depicts the OpenStack/Ceph technology stack.



## Important

To use Ceph Block Devices with OpenStack, you must have access to a running Ceph Storage Cluster.

Three parts of OpenStack integrate with Ceph's block devices:

- **Images:** OpenStack Glance manages images for VMs. Images are immutable. OpenStack treats images as binary blobs and downloads them accordingly.
- **Volumes:** Volumes are block devices. OpenStack uses volumes to boot VMs, or to attach volumes to running VMs. OpenStack manages volumes using Cinder services.
- **Guest Disks:** Guest disks are guest operating system disks. By default, when you boot a virtual machine, its disk appears as a file on the file system of the hypervisor (usually under `/var/lib/nova/instances/<uuid>/`). Prior to OpenStack Havana, the only way to boot a VM in Ceph was to use the boot-from-volume functionality of Cinder. However, now it is possible to boot every virtual machine inside Ceph directly without using Cinder, which is advantageous because it allows you to perform maintenance operations easily with the live-migration process.

Additionally, if your hypervisor dies it is also convenient to trigger `nova evacuate` and reinstate the virtual machine elsewhere almost seamlessly. In doing so, `exclusive locks` prevent multiple compute nodes from concurrently accessing the guest disk.

You can use OpenStack Glance to store images as Ceph Block Devices, and you can use Cinder to boot a VM using a copy-on-write clone of an image.

The instructions below detail the setup for Glance, Cinder and Nova, although they do not have to be used together. You may store images in Ceph block devices while running VMs using a local disk, or vice versa.

#### Important

Using QCOW2 for hosting a virtual machine disk is NOT recommended. If you want to boot virtual machines in Ceph (ephemeral backend or boot from volume), please use the `raw` image format within Glance.

## Create a Pool

By default, Ceph block devices live within the `rbd` pool. You may use any suitable pool by specifying it explicitly. We recommend creating a pool for Cinder and a pool for Glance. Ensure your Ceph cluster is running, then create the pools.

1. `ceph osd pool create volumes`
2. `ceph osd pool create images`
3. `ceph osd pool create backups`
4. `ceph osd pool create vms`

See [Create a Pool](#) for detail on specifying the number of placement groups for your pools, and [Placement Groups](#) for details on the number of placement groups you should set for your pools.

Newly created pools must be initialized prior to use. Use the `rbd` tool to initialize the pools:

1. `rbd pool init volumes`
2. `rbd pool init images`
3. `rbd pool init backups`
4. `rbd pool init vms`

## Configure OpenStack Ceph Clients

The nodes running `glance-api`, `cinder-volume`, `nova-compute` and `cinder-backup` act as Ceph clients. Each requires the `ceph.conf` file:

1. `ssh {your-openstack-server} sudo tee /etc/ceph/ceph.conf </etc/ceph/ceph.conf`

# Install Ceph client packages

On the `glance-api` node, you will need the Python bindings for `librbd` :

1. `sudo apt-get install python-rbd`
2. `sudo yum install python-rbd`

On the `nova-compute`, `cinder-backup` and on the `cinder-volume` node, use both the Python bindings and the client command line tools:

1. `sudo apt-get install ceph-common`
2. `sudo yum install ceph-common`

## Setup Ceph Client Authentication

If you have `cephx authentication` enabled, create a new user for Nova/Cinder and Glance. Execute the following:

- ```
ceph auth get-or-create client.glance mon 'profile rbd' osd 'profile rbd pool=images' mgr 'profile rbd
1. pool=images'
    ceph auth get-or-create client.cinder mon 'profile rbd' osd 'profile rbd pool=volumes, profile rbd pool=vms,
2. profile rbd-read-only pool=images' mgr 'profile rbd pool=volumes, profile rbd pool=vms'
    ceph auth get-or-create client.cinder-backup mon 'profile rbd' osd 'profile rbd pool=backups' mgr 'profile rbd
3. pool=backups'
```

Add the keyrings for `client.cinder`, `client.glance`, and `client.cinder-backup` to the appropriate nodes and change their ownership:

- ```
ceph auth get-or-create client.glance | ssh {your-glance-api-server} sudo tee
1. /etc/ceph/ceph.client.glance.keyring
2. ssh {your-glance-api-server} sudo chown glance:glance /etc/ceph/ceph.client.glance.keyring
3. ceph auth get-or-create client.cinder | ssh {your-volume-server} sudo tee /etc/ceph/ceph.client.cinder.keyring
4. ssh {your-cinder-volume-server} sudo chown cinder:cinder /etc/ceph/ceph.client.cinder.keyring
    ceph auth get-or-create client.cinder-backup | ssh {your-cinder-backup-server} sudo tee
5. /etc/ceph/ceph.client.cinder-backup.keyring
6. ssh {your-cinder-backup-server} sudo chown cinder:cinder /etc/ceph/ceph.client.cinder-backup.keyring
```

Nodes running `nova-compute` need the keyring file for the `nova-compute` process:

- ```
ceph auth get-or-create client.cinder | ssh {your-nova-compute-server} sudo tee
1. /etc/ceph/ceph.client.cinder.keyring
```

They also need to store the secret key of the `client.cinder` user in `libvirt`. The `libvirt` process needs it to access the cluster while attaching a block device from Cinder.

Create a temporary copy of the secret key on the nodes running `nova-compute` :

```
1. ceph auth get-key client.cinder | ssh {your-compute-node} tee client.cinder.key
```

Then, on the compute nodes, add the secret key to `libvirt` and remove the temporary copy of the key:

```
1. uuidgen
2. 457eb676-33da-42ec-9a8c-9293d545c337
3.
4. cat > secret.xml <<EOF
5. <secret ephemeral='no' private='no'>
6.   <uuid>457eb676-33da-42ec-9a8c-9293d545c337</uuid>
7.   <usage type='ceph'>
8.     <name>client.cinder secret</name>
9.   </usage>
10.  </secret>
11. EOF
12. sudo virsh secret-define --file secret.xml
13. Secret 457eb676-33da-42ec-9a8c-9293d545c337 created
    sudo virsh secret-set-value --secret 457eb676-33da-42ec-9a8c-9293d545c337 --base64 $(cat client.cinder.key) &&
14. rm client.cinder.key secret.xml
```

Save the uuid of the secret for configuring `nova-compute` later.

### Important

You don't necessarily need the UUID on all the compute nodes. However from a platform consistency perspective, it's better to keep the same UUID.

## Configure OpenStack to use Ceph

### Configuring Glance

Glance can use multiple back ends to store images. To use Ceph block devices by default, configure Glance like the following.

### Kilo and after

Edit `/etc/glance/glance-api.conf` and add under the `[glance_store]` section:

```
1. [glance_store]
2. stores = rbd
3. default_store = rbd
4. rbd_store_pool = images
5. rbd_store_user = glance
6. rbd_store_ceph_conf = /etc/ceph/ceph.conf
7. rbd_store_chunk_size = 8
```

For more information about the configuration options available in Glance please refer

to the OpenStack Configuration Reference: <http://docs.openstack.org/>.

## Enable copy-on-write cloning of images

Note that this exposes the back end location via Glance's API, so the endpoint with this option enabled should not be publicly accessible.

Any OpenStack version except Mitaka

If you want to enable copy-on-write cloning of images, also add under the `[DEFAULT]` section:

```
1. show_image_direct_url = True
```

## Disable cache management (any OpenStack version)

Disable the Glance cache management to avoid images getting cached under

```
/var/lib/glance/image-cache/ , assuming your configuration file has flavor = keystone+cachemanagement :
```

```
1. [paste_deploy]
2. flavor = keystone
```

## Image properties

We recommend to use the following properties for your images:

- `hw_scsi_model=virtio-scsi` : add the virtio-scsi controller and get better performance and support for discard operation
- `hw_disk_bus=scsi` : connect every cinder block devices to that controller
- `hw_qemu_guest_agent=yes` : enable the QEMU guest agent
- `os_require_quiesce=yes` : send fs-freeze/thaw calls through the QEMU guest agent

## Configuring Cinder

OpenStack requires a driver to interact with Ceph block devices. You must also specify the pool name for the block device. On your OpenStack node, edit `/etc/cinder/cinder.conf` by adding:

```
1. [DEFAULT]
2. ...
3. enabled_backends = ceph
4. glance_api_version = 2
5. ...
6. [ceph]
7. volume_driver = cinder.volume.drivers.rbd.RBDDriver
```

```

8. volume_backend_name = ceph
9. rbd_pool = volumes
10. rbd_ceph_conf = /etc/ceph/ceph.conf
11. rbd_flatten_volume_from_snapshot = false
12. rbd_max_clone_depth = 5
13. rbd_store_chunk_size = 4
14. rados_connect_timeout = -1

```

If you are using [cephx authentication](#), also configure the user and uuid of the secret you added to `libvirt` as documented earlier:

```

1. [ceph]
2. ...
3. rbd_user = cinder
4. rbd_secret_uuid = 457eb676-33da-42ec-9a8c-9293d545c337

```

Note that if you are configuring multiple cinder back ends, `glance_api_version = 2` must be in the `[DEFAULT]` section.

## Configuring Cinder Backup

OpenStack Cinder Backup requires a specific daemon so don't forget to install it. On your Cinder Backup node, edit `/etc/cinder/cinder.conf` and add:

```

1. backup_driver = cinder.backup.drivers.ceph
2. backup_ceph_conf = /etc/ceph/ceph.conf
3. backup_ceph_user = cinder-backup
4. backup_ceph_chunk_size = 134217728
5. backup_ceph_pool = backups
6. backup_ceph_stripe_unit = 0
7. backup_ceph_stripe_count = 0
8. restore_discard_excess_bytes = true

```

## Configuring Nova to attach Ceph RBD block device

In order to attach Cinder devices (either normal block or by issuing a boot from volume), you must tell Nova (and libvirt) which user and UUID to refer to when attaching the device. libvirt will refer to this user when connecting and authenticating with the Ceph cluster.

```

1. [libvirt]
2. ...
3. rbd_user = cinder
4. rbd_secret_uuid = 457eb676-33da-42ec-9a8c-9293d545c337

```

These two flags are also used by the Nova ephemeral back end.

## Configuring Nova

In order to boot virtual machines directly from Ceph volumes, you must configure the ephemeral backend for Nova.

It is recommended to enable the RBD cache in your Ceph configuration file; this has been enabled by default since the Giant release. Moreover, enabling the client admin socket allows the collection of metrics and can be invaluable for troubleshooting.

This socket can be accessed on the hypervisor (Nova compute) node:

```
1. ceph daemon /var/run/ceph/ceph-client.cinder.19195.32310016.asok help
```

To enable RBD cache and admin sockets, ensure that on each hypervisor's `ceph.conf` contains:

```
1. [client]
2.   rbd cache = true
3.   rbd cache writethrough until flush = true
4.   admin socket = /var/run/ceph/guests/$cluster-$type.$id.$pid.$cctid.asok
5.   log file = /var/log/qemu/qemu-guest-$pid.log
6.   rbd concurrent management ops = 20
```

Configure permissions for these directories:

```
1. mkdir -p /var/run/ceph/guests/ /var/log/qemu/
2. chown qemu:libvirtd /var/run/ceph/guests /var/log/qemu/
```

Note that user `qemu` and group `libvirtd` can vary depending on your system. The provided example works for RedHat based systems.

Tip

If your virtual machine is already running you can simply restart it to enable the admin socket

## Restart OpenStack

To activate the Ceph block device driver and load the block device pool name into the configuration, you must restart the related OpenStack services. For Debian based systems execute these commands on the appropriate nodes:

```
1. sudo glance-control api restart
2. sudo service nova-compute restart
3. sudo service cinder-volume restart
4. sudo service cinder-backup restart
```

For Red Hat based systems execute:

1. sudo service openstack-glance-api restart
2. sudo service openstack-nova-compute restart
3. sudo service openstack-cinder-volume restart
4. sudo service openstack-cinder-backup restart

Once OpenStack is up and running, you should be able to create a volume and boot from it.

## Booting from a Block Device

You can create a volume from an image using the Cinder command line tool:

1. cinder create --image-id {id of image} --display-name {name of volume} {size of volume}

You can use `qemu-img` to convert from one format to another. For example:

1. qemu-img convert -f {source-format} -O {output-format} {source-filename} {output-filename}
2. qemu-img convert -f qcow2 -O raw precise-cloudimg.img precise-cloudimg.raw

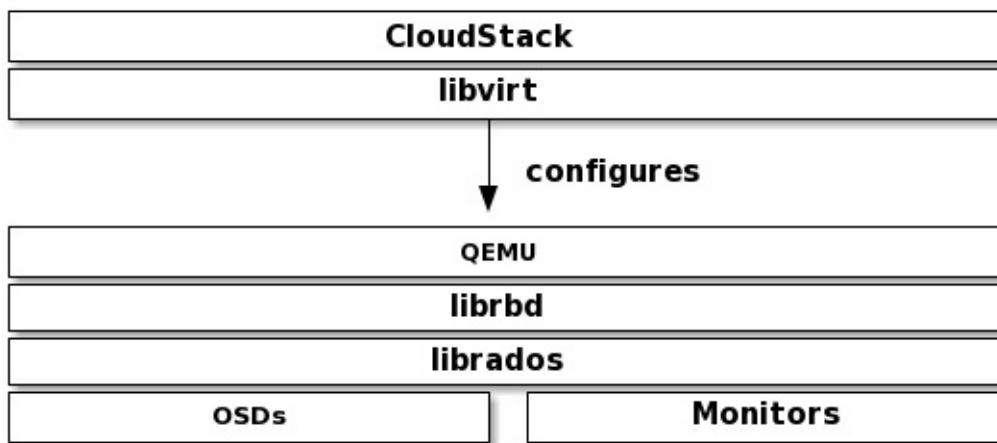
When Glance and Cinder are both using Ceph block devices, the image is a copy-on-write clone, so new volumes are created quickly. In the OpenStack dashboard, you can boot from that volume by performing the following steps:

1. Launch a new instance.
2. Choose the image associated to the copy-on-write clone.
3. Select 'boot from volume'.
4. Select the volume you created.

# Block Devices and CloudStack

You may use Ceph Block Device images with CloudStack 4.0 and higher through `libvirt`, which configures the QEMU interface to `librbd`. Ceph stripes block device images as objects across the cluster, which means that large Ceph Block Device images have better performance than a standalone server!

To use Ceph Block Devices with CloudStack 4.0 and higher, you must install QEMU, `libvirt`, and CloudStack first. We recommend using a separate physical host for your CloudStack installation. CloudStack recommends a minimum of 4GB of RAM and a dual-core processor, but more CPU and RAM will perform better. The following diagram depicts the CloudStack/Ceph technology stack.



## Important

To use Ceph Block Devices with CloudStack, you must have access to a running Ceph Storage Cluster.

CloudStack integrates with Ceph's block devices to provide CloudStack with a back end for CloudStack's Primary Storage. The instructions below detail the setup for CloudStack Primary Storage.

## Note

We recommend installing with Ubuntu 14.04 or later so that you can use package installation instead of having to compile `libvirt` from source.

Installing and configuring QEMU for use with CloudStack doesn't require any special handling. Ensure that you have a running Ceph Storage Cluster. Install QEMU and configure it for use with Ceph; then, install `libvirt` version 0.9.13 or higher (you may need to compile from source) and ensure it is running with Ceph.

## Note

Ubuntu 14.04 and CentOS 7.2 will have `libvirt` with RBD storage pool support enabled by default.

## Create a Pool

By default, Ceph block devices use the `rbd` pool. Create a pool for CloudStack NFS Primary Storage. Ensure your Ceph cluster is running, then create the pool.

```
1. ceph osd pool create cloudstack
```

See [Create a Pool](#) for details on specifying the number of placement groups for your pools, and [Placement Groups](#) for details on the number of placement groups you should set for your pools.

A newly created pool must be initialized prior to use. Use the `rbd` tool to initialize the pool:

```
1. rbd pool init cloudstack
```

## Create a Ceph User

To access the Ceph cluster we require a Ceph user which has the correct credentials to access the `cloudstack` pool we just created. Although we could use `client.admin` for this, it's recommended to create a user with only access to the `cloudstack` pool.

```
1. ceph auth get-or-create client.cloudstack mon 'profile rbd' osd 'profile rbd pool=cloudstack'
```

Use the information returned by the command in the next step when adding the Primary Storage.

See [User Management](#) for additional details.

## Add Primary Storage

To add a Ceph block device as Primary Storage, the steps include:

1. Log in to the CloudStack UI.
2. Click **Infrastructure** on the left side navigation bar.
3. Select **View All** under **Primary Storage**.
4. Click the **Add Primary Storage** button on the top right hand side.
5. Fill in the following information, according to your infrastructure setup:

- Scope (i.e. Cluster or Zone-Wide).
  - Zone.
  - Pod.
  - Cluster.
  - Name of Primary Storage.
- For **Protocol**, select `RBD`.
  - For **Provider**, select the appropriate provider type (i.e. DefaultPrimary, SolidFire, SolidFireShared, or CloudByte). Depending on the provider chosen, fill out the information pertinent to your setup.
6. Add cluster information (`cephx` is supported).
- For **RADOS Monitor**, provide the IP address of a Ceph monitor node.
  - For **RADOS Pool**, provide the name of an RBD pool.
  - For **RADOS User**, provide a user that has sufficient rights to the RBD pool.  
Note: Do not include the `client.` part of the user.
  - For **RADOS Secret**, provide the secret the user's secret.
  - **Storage Tags** are optional. Use tags at your own discretion. For more information about storage tags in CloudStack, refer to [Storage Tags](#).
7. Click **OK**.

## Create a Disk Offering

---

To create a new disk offering, refer to [Create a New Disk Offering](#). Create a disk offering so that it matches the `rbd` tag. The `storagePoolAllocator` will choose the `rbd` pool when searching for a suitable storage pool. If the disk offering doesn't match the `rbd` tag, the `storagePoolAllocator` may select the pool you created (e.g., `cloudstack` ).

## Limitations

---

- CloudStack will only bind to one monitor (You can however create a Round Robin DNS record over multiple monitors)

# Ceph iSCSI Gateway

The iSCSI Gateway presents a Highly Available (HA) iSCSI target that exports RADOS Block Device (RBD) images as SCSI disks. The iSCSI protocol allows clients (initiators) to send SCSI commands to storage devices (targets) over a TCP/IP network, enabling clients without native Ceph client support to access Ceph block storage. These include Microsoft Windows and even BIOS.

Each iSCSI gateway exploits the Linux IO target kernel subsystem (LIO) to provide iSCSI protocol support. LIO utilizes userspace passthrough (TCMU) to interact with Ceph's librbd library and expose RBD images to iSCSI clients. With Ceph's iSCSI gateway you can provision a fully integrated block-storage infrastructure with all the features and benefits of a conventional Storage Area Network (SAN).

- Requirements
- Configuring the iSCSI Target
- Configuring the iSCSI Initiators
- Monitoring the iSCSI Gateways

# iSCSI Gateway Requirements

It is recommended to provision two to four iSCSI gateway nodes to realize a highly available Ceph iSCSI gateway solution.

For hardware recommendations, see [Hardware Recommendations](#).

## Note

On iSCSI gateway nodes the memory footprint is a function of the RBD images mapped and can grow to be large. Plan memory requirements accordingly based on the number RBD images to be mapped.

There are no specific iSCSI gateway options for the Ceph Monitors or OSDs, but it is important to lower the default heartbeat interval for detecting down OSDs to reduce the possibility of initiator timeouts. The following configuration options are suggested:

1. [osd]
2. osd heartbeat grace = 20
3. osd heartbeat interval = 5

- Updating Running State From a Ceph Monitor Node

- ```
1. ceph tell <daemon_type>.<id> config set <parameter_name> <new_value>

1. ceph tell osd.* config set osd_heartbeat_grace 20
2. ceph tell osd.* config set osd_heartbeat_interval 5
```

- Updating Running State On Each OSD Node

- ```
1. ceph daemon <daemon_type>.<id> config set osd_client_watch_timeout 15

1. ceph daemon osd.0 config set osd_heartbeat_grace 20
2. ceph daemon osd.0 config set osd_heartbeat_interval 5
```

For more details on setting Ceph's configuration options, see [Configuring Ceph](#). Be sure to persist these settings in `/etc/ceph.conf` or, on Mimic and later releases, in the centralized config store.

# iSCSI Targets

---

Traditionally, block-level access to a Ceph storage cluster has been limited to QEMU and `librbd`, which is a key enabler for adoption within OpenStack environments. Starting with the Ceph Luminous release, block-level access is expanding to offer standard iSCSI support allowing wider platform usage, and potentially opening new use cases.

- Red Hat Enterprise Linux/CentOS 7.5 (or newer); Linux kernel v4.16 (or newer)
- A working Ceph Storage cluster, deployed with `ceph-ansible` or using the command-line interface
- iSCSI gateways nodes, which can either be colocated with OSD nodes or on dedicated nodes
- Separate network subnets for iSCSI front-end traffic and Ceph back-end traffic

A choice of using Ansible or the command-line interface are the available deployment methods for installing and configuring the Ceph iSCSI gateway:

- [Using Ansible](#)
- [Using the Command Line Interface](#)

# Configuring the iSCSI Target using Ansible

The Ceph iSCSI gateway is the iSCSI target node and also a Ceph client node. The Ceph iSCSI gateway can be provisioned on dedicated node or be colocated on a Ceph Object Store Disk (OSD) node. The following steps will install and configure the Ceph iSCSI gateway for basic operation.

## Requirements:

- A running Ceph Luminous (12.2.x) cluster or newer
- Red Hat Enterprise Linux/CentOS 7.5 (or newer); Linux kernel v4.16 (or newer)
- The `ceph-iscsi` package installed on all the iSCSI gateway nodes

## Installation:

1. On the Ansible installer node, which could be either the administration node or a dedicated deployment node, perform the following steps:

- i. As `root`, install the `ceph-ansible` package:

```
1. # yum install ceph-ansible
```

- ii. Add an entry in `/etc/ansible/hosts` file for the gateway group:

```
1. [iscsigws]
2. ceph-igw-1
3. ceph-igw-2
```

## Note

If co-locating the iSCSI gateway with an OSD node, then add the OSD node to the `[iscsigws]` section.

## Configuration:

The `ceph-ansible` package places a file in the `/usr/share/ceph-ansible/group_vars/` directory called `iscsigws.yml.sample`. Create a copy of this sample file named `iscsigws.yml`. Review the following Ansible variables and descriptions, and update accordingly. See the `iscsigws.yml.sample` for a full list of advanced variables.

Variable	Meaning/Purpose
<code>seed_monitor</code>	Each gateway needs access to the ceph cluster for rados and rbd calls. This means the iSCSI gateway must have an appropriate <code>/etc/ceph/</code> directory defined. The <code>seed_monitor</code> host is used to populate the iSCSI gateway's <code>/etc/ceph/</code> directory.

<code>cluster_name</code>	Define a custom storage cluster name.
<code>gateway_keyring</code>	Define a custom keyring name.
<code>deploy_settings</code>	If set to <code>true</code> , then deploy the settings when the playbook is ran.
<code>perform_system_checks</code>	This is a boolean value that checks for multipath and lvm configuration settings on each gateway. It must be set to true for at least the first run to ensure multipathd and lvm are configured properly.
<code>api_user</code>	The user name for the API. The default is <code>admin</code> .
<code>api_password</code>	The password for using the API. The default is <code>admin</code> .
<code>api_port</code>	The TCP port number for using the API. The default is <code>5000</code> .
<code>api_secure</code>	True if TLS must be used. The default is <code>false</code> . If true the user must create the necessary certificate and key files. See the gwcli man file for details.
<code>trusted_ip_list</code>	A list of IPv4 or IPv6 addresses who have access to the API. By default, only the iSCSI gateway nodes have access.

## Deployment:

Perform the followint steps on the Ansible installer node.

- As `root`, execute the Ansible playbook:

```
1. # cd /usr/share/ceph-ansible
2. # ansible-playbook site.yml --limit iscsigw
```

### Note

The Ansible playbook will handle RPM dependencies, setting up daemons, and installing gwcli so it can be used to create iSCSI targets and export RBD images as LUNs. In past versions, `iscsigws.yml` could define the iSCSI target and other objects like clients, images and LUNs, but this is no longer supported.

- Verify the configuration from an iSCSI gateway node:

```
1. # gwcli ls
```

### Note

See the [Configuring the iSCSI Target using the Command Line Interface](#) section to

create gateways, LUNs, and clients using the gwcli tool.

## Important

Attempting to use the `targetcli` tool to change the configuration will cause problems including ALUA misconfiguration and path failover issues. There is the potential to corrupt data, to have mismatched configuration across iSCSI gateways, and to have mismatched WWN information, leading to client multipath problems.

## Service Management:

The `ceph-iscsi` package installs the configuration management logic and a Systemd service called `rbd-target-api`. When the Systemd service is enabled, the `rbd-target-api` will start at boot time and will restore the Linux IO state. The Ansible playbook disables the target service during the deployment. Below are the outcomes of when interacting with the `rbd-target-api` Systemd service.

```
1. # systemctl <start|stop|restart|reload> rbd-target-api
```

- `reload`

A reload request will force `rbd-target-api` to reread the configuration and apply it to the current running environment. This is normally not required, since changes are deployed in parallel from Ansible to all iSCSI gateway nodes

- `stop`

A stop request will close the gateway's portal interfaces, dropping connections to clients and wipe the current LIO configuration from the kernel. This returns the iSCSI gateway to a clean state. When clients are disconnected, active I/O is rescheduled to the other iSCSI gateways by the client side multipathing layer.

## Removing the Configuration:

The `ceph-ansible` package provides an Ansible playbook to remove the iSCSI gateway configuration and related RBD images. The Ansible playbook is `/usr/share/ceph-ansible/purge_gateways.yml`. When this Ansible playbook is ran a prompted for the type of purge to perform:

`lio` :

In this mode the LIO configuration is purged on all iSCSI gateways that are defined. Disks that were created are left untouched within the Ceph storage cluster.

`all` :

When `all` is chosen, the LIO configuration is removed together with `all` RBD images that were defined within the iSCSI gateway environment, other unrelated RBD images will not be removed. Ensure the correct mode is chosen, this operation will delete

data.

## Warning

A purge operation is destructive action against your iSCSI gateway environment.

## Warning

A purge operation will fail, if RBD images have snapshots or clones and are exported through the Ceph iSCSI gateway.

```
1. [root@rh7-iscsi-client ceph-ansible]# ansible-playbook purge_gateways.yml
2. Which configuration elements should be purged? (all, lio or abort) [abort]: all
3.
4.
5. PLAY [Confirm removal of the iSCSI gateway configuration] ****
6.
7.
8. GATHERING FACTS ****
9. ok: [localhost]
10.
11.
12. TASK: [Exit playbook if user aborted the purge] ****
13. skipping: [localhost]
14.
15.
16. TASK: [set_fact ] ****
17. ok: [localhost]
18.
19.
20. PLAY [Removing the gateway configuration] ****
21.
22.
23. GATHERING FACTS ****
24. ok: [ceph-igw-1]
25. ok: [ceph-igw-2]
26.
27.
28. TASK: [igw_purge | purging the gateway configuration] ****
29. changed: [ceph-igw-1]
30. changed: [ceph-igw-2]
31.
32.
33. TASK: [igw_purge | deleting configured rbd devices] ****
34. changed: [ceph-igw-1]
35. changed: [ceph-igw-2]
36.
37.
38. PLAY RECAP ****
39. ceph-igw-1 : ok=3    changed=2    unreachable=0    failed=0
40. ceph-igw-2 : ok=3    changed=2    unreachable=0    failed=0
41. localhost   : ok=2    changed=0    unreachable=0    failed=0
```

# Configuring the iSCSI Target using the Command Line Interface

The Ceph iSCSI gateway is both an iSCSI target and a Ceph client; think of it as a “translator” between Ceph’s RBD interface and the iSCSI standard. The Ceph iSCSI gateway can run on a standalone node or be colocated with other daemons eg. on a Ceph Object Store Disk (OSD) node. When co-locating, ensure that sufficient CPU and memory are available to share. The following steps install and configure the Ceph iSCSI gateway for basic operation.

## Requirements:

- A running Ceph Luminous or later storage cluster
- Red Hat Enterprise Linux/CentOS 7.5 (or newer); Linux kernel v4.16 (or newer)
- The following packages must be installed from your Linux distribution’s software repository:
  - `targetcli-2.1.fb47` or newer package
  - `python-rtslib-2.1.fb68` or newer package
  - `tcmu-runner-1.4.0` or newer package
  - `ceph-iscsi-3.2` or newer package

### Important

If previous versions of these packages exist, then they must be removed first before installing the newer versions.

Do the following steps on the Ceph iSCSI gateway node before proceeding to the *Installing* section:

1. If the Ceph iSCSI gateway is not colocated on an OSD node, then copy the Ceph configuration files, located in `/etc/ceph/`, from a running Ceph node in the storage cluster to the iSCSI Gateway node. The Ceph configuration files must exist on the iSCSI gateway node under `/etc/ceph/`.
2. Install and configure the [Ceph Command-line Interface](#)
3. If needed, open TCP ports 3260 and 5000 on the firewall.

### Note

Access to port 5000 should be restricted to a trusted internal network or only the individual hosts where `gwcli` is used or `ceph-mgr` daemons are running.

#### 4. Create a new or use an existing RADOS Block Device (RBD).

##### Installing:

If you are using the upstream ceph-iscsi package follow the [manual install instructions](#).

For rpm based instructions execute the following commands:

1. As `root`, on all iSCSI gateway nodes, install the `ceph-iscsi` package:

```
1. # yum install ceph-iscsi
```

2. As `root`, on all iSCSI gateway nodes, install the `tcmu-runner` package:

```
1. # yum install tcmu-runner
```

##### Setup:

1. gwcli requires a pool with the name `rbd`, so it can store metadata like the iSCSI configuration. To check if this pool has been created run:

```
1. # ceph osd lspools
```

If it does not exist instructions for creating pools can be found on the [RADOS pool operations page](#).

2. As `root`, on a iSCSI gateway node, create a file named `iscsi-gateway.cfg` in the `/etc/ceph/` directory:

```
1. # touch /etc/ceph/iscsi-gateway.cfg
```

- i. Edit the `iscsi-gateway.cfg` file and add the following lines:

```
``` [config]
```

Name of the Ceph storage cluster. A suitable Ceph configuration file allowing

access to the Ceph storage cluster from the gateway node is required, if not

colocated on an OSD node.

```
cluster_name = ceph
```

Place a copy of the ceph cluster's admin keyring in the gateway's /etc/ceph

directory and reference the filename here

```
gateway_keyring = ceph.client.admin.keyring
```

```

1. # API settings.
2. # The API supports a number of options that allow you to tailor it to your
3. # local environment. If you want to run the API under https, you will need to
4. # create cert/key files that are compatible for each iSCSI gateway node, that is
5. # not locked to a specific node. SSL cert and key files *must* be called
6. # 'iscsi-gateway.crt' and 'iscsi-gateway.key' and placed in the '/etc/ceph/' directory
7. # on *each* gateway node. With the SSL files in place, you can use 'api_secure = true'
8. # to switch to https mode.
9.
10. # To support the API, the bare minimum settings are:
11. api_secure = false
12.
13. # Additional API configuration options are as follows, defaults shown.
14. # api_user = admin
15. # api_password = admin
16. # api_port = 5001
17. # trusted_ip_list = 192.168.0.10,192.168.0.11
18. ``
19.
20. Note
21.
22. trusted\_\_ip\_\_list is a list of IP addresses on each iSCSI gateway that will be used for management
   operations like target creation, LUN exporting, etc. The IP can be the same that will be used for iSCSI data,
23. like READ/WRITE commands to/from the RBD image, but using separate IPs is recommended.
24.
25. Important
26.
27. The `iscsi-gateway.cfg` file must be identical on all iSCSI gateway nodes.
28. 2. As `root`, copy the `iscsi-gateway.cfg` file to all iSCSI gateway nodes.
```

- As `root`, on all iSCSI gateway nodes, enable and start the API service:

```

1. # systemctl daemon-reload
2.
3. # systemctl enable rbd-target-gw
4. # systemctl start rbd-target-gw
5.
6. # systemctl enable rbd-target-api
7. # systemctl start rbd-target-api

```

## Configuring:

gwcli will create and configure the iSCSI target and RBD images and copy the configuration across the gateways setup in the last section. Lower level tools including targetcli and rbd can be used to query the local configuration, but should not be used to modify it. This next section will demonstrate how to create a iSCSI target and export a RBD image as LUN 0.

- As `root`, on a iSCSI gateway node, start the iSCSI gateway command-line interface:

```
1. # gwcli
```

- Go to `iscsi-targets` and create a target with the name `iqn.2003-01.com.redhat.iscsi-gw:iscsi-igw`:

```

1. > /> cd /iscsi-target
2. > /iscsi-target> create iqn.2003-01.com.redhat.iscsi-gw:iscsi-igw

```

- Create the iSCSI gateways. The IPs used below are the ones that will be used for iSCSI data like READ and WRITE commands. They can be the same IPs used for management operations listed in `trusted_ip_list`, but it is recommended that different IPs are used.

```

1. > /iscsi-target> cd iqn.2003-01.com.redhat.iscsi-gw:iscsi-igw/gateways
2. > /iscsi-target...-igw/gateways> create ceph-gw-1 10.172.19.21
3. > /iscsi-target...-igw/gateways> create ceph-gw-2 10.172.19.22

```

If not using RHEL/CentOS or using an upstream or `ceph-iscsi-test` kernel, the `skipchecks=true` argument must be used. This will avoid the Red Hat kernel and rpm checks:

```

1. > /iscsi-target> cd iqn.2003-01.com.redhat.iscsi-gw:iscsi-igw/gateways
2. > /iscsi-target...-igw/gateways> create ceph-gw-1 10.172.19.21 skipchecks=true
3. > /iscsi-target...-igw/gateways> create ceph-gw-2 10.172.19.22 skipchecks=true

```

- Add a RBD image with the name `disk_1` in the pool `rbd`:

```
1. > /iscsi-target...-igw/gateways> cd /disks
```

```
2. > /disks> create pool=rbd image=disk_1 size=90G
```

5. Create a client with the initiator name iqn.1994-05.com.redhat:rh7-client:

```
1. > /disks> cd /iscsi-target/iqn.2003-01.com.redhat.iscsi-gw:iscsi-igw/hosts  
2. > /iscsi-target...eph-igw/hosts> create iqn.1994-05.com.redhat:rh7-client
```

6. Set the client's CHAP username to myiscsiusername and password to myiscsipassword:

```
1. > /iscsi-target...at:rh7-client> auth username=myiscsiusername password=myiscsipassword
```

#### Warning

CHAP must always be configured. Without CHAP, the target will reject any login requests.

7. Add the disk to the client:

```
1. > /iscsi-target...at:rh7-client> disk add rbd/disk_1
```

The next step is to configure the iSCSI initiators.

# Configuring the iSCSI Initiators

- [iSCSI Initiator for Linux](#)
- [iSCSI Initiator for Microsoft Windows](#)
- [iSCSI Initiator for VMware ESX](#)

Warning

Applications that use SCSI persistent group reservations (PGR) and SCSI 2 based reservations are not supported when exporting a RBD image through more than one iSCSI gateway.

# Monitoring Ceph iSCSI gateways

Ceph provides a tool for iSCSI gateway environments to monitor performance of exported RADOS Block Device (RBD) images.

The `gwtop` tool is a `top`-like tool that displays aggregated performance metrics of RBD images that are exported to clients over iSCSI. The metrics are sourced from a Performance Metrics Domain Agent (PMDA). Information from the Linux-IO target (LIO) PMDA is used to list each exported RBD image, the connected client, and its associated I/O metrics.

## Requirements:

- A running Ceph iSCSI gateway

## Installing:

- As `root`, install the `ceph-iscsi-tools` package on each iSCSI gateway node:

```
1. # yum install ceph-iscsi-tools
```

- As `root`, install the performance co-pilot package on each iSCSI gateway node:

```
1. # yum install pcp
```

- As `root`, install the LIO PMDA package on each iSCSI gateway node:

```
1. # yum install pcp-pmda-lio
```

- As `root`, enable and start the performance co-pilot service on each iSCSI gateway node:

```
1. # systemctl enable pmcd
2. # systemctl start pmcd
```

- As `root`, register the `pcp-pmda-lio` agent:

```
1. cd /var/lib/pcp/pmdas/lio
2. ./Install
```

By default, `gwtop` assumes the iSCSI gateway configuration object is stored in a RADOS object called `gateway.conf` in the `rbd` pool. This configuration defines the iSCSI gateways to contact for gathering the performance statistics. This can be overridden by using either the `-g` or `-c` flags. See `gwtop --help` for more details.

The LIO configuration determines which type of performance statistics to extract from performance co-pilot. When `gwtop` starts it looks at the LIO configuration, and if it finds user-space disks, then `gwtop` selects the LIO collector automatically.

### Example ``gwtop`` Outputs

1.	<code>gwtop</code>	2/2 Gateways	CPU%	MIN: 4	MAX: 5	Network	Total In:	2M	Out:	3M	10:20:00	
2.	Capacity:	8G	Disks:	8	IOPS:	503	Clients:	1	Ceph:	HEALTH_OK	OSDs:	3
3.	Pool	Image	Src	Size	iops	rMB/s	wMB/s	Client				
4.	iscsi	t1703		500M	0	0.00	0.00					
5.	iscsi	testme1		500M	0	0.00	0.00					
6.	iscsi	testme2		500M	0	0.00	0.00					
7.	iscsi	testme3		500M	0	0.00	0.00					
8.	iscsi	testme5		500M	0	0.00	0.00					
9.	rbd	myhost_1	T	4G	504	1.95	0.00	rh460p	(CON)			
10.	rbd	test_2		1G	0	0.00	0.00					
11.	rbd	testme		500M	0	0.00	0.00					

In the *Client* column, `(CON)` means the iSCSI initiator (client) is currently logged into the iSCSI gateway. If `-multi-` is displayed, then multiple clients are mapped to the single RBD image.

# rbd – manage rados block device (RBD) images

## Synopsis

```
rbd [ -c ceph.conf ] [ -m monaddr ] [-cluster cluster-name] [ -p | --pool pool ] [ command ... ]
```

## Description

**rbd** is a utility for manipulating rados block device (RBD) images, used by the Linux rbd driver and the rbd storage driver for QEMU/KVM. RBD images are simple block devices that are striped over objects and stored in a RADOS object store. The size of the objects the image is striped over must be a power of two.

## Options

```
-c ceph.conf``, --conf ceph.conf
```

Use `ceph.conf` configuration file instead of the default `/etc/ceph/ceph.conf` to determine monitor addresses during startup.

```
-m monaddress[:port]
```

Connect to specified monitor (instead of looking through `ceph.conf`).

```
--cluster cluster-name
```

Use different cluster name as compared to default cluster name `ceph`.

```
-p pool-name``, --pool pool-name
```

Interact with the given pool. Required by most commands.

```
--namespace namespace-name
```

Use a pre-defined image namespace within a pool

```
--no-progress
```

Do not output progress information (goes to standard error by default for some commands).

## Parameters

--image-format    format-id

Specifies which object layout to use. The default is 2.

- format 1 - (deprecated) Use the original format for a new rbd image. This format is understood by all versions of librbd and the kernel rbd module, but does not support newer features like cloning.
- format 2 - Use the second rbd format, which is supported by librbd since the Bobtail release and the kernel rbd module since kernel 3.10 (except for “fancy” striping, which is supported since kernel 4.17). This adds support for cloning and is more easily extensible to allow more features in the future.

-s    size-in-M/G/T` ,    --size    size-in-M/G/T

Specifies the size of the new rbd image or the new size of the existing rbd image in M/G/T. If no suffix is given, unit M is assumed.

--object-size    size-in-B/K/M

Specifies the object size in B/K/M. Object size will be rounded up the nearest power of two; if no suffix is given, unit B is assumed. The default object size is 4M, smallest is 4K and maximum is 32M.

--stripe-unit    size-in-B/K/M

Specifies the stripe unit size in B/K/M. If no suffix is given, unit B is assumed. See striping section (below) for more details.

--stripe-count    num

Specifies the number of objects to stripe over before looping back to the first object. See striping section (below) for more details.

--snap    snap

Specifies the snapshot name for the specific operation.

--id    username

Specifies the username (without the `client.` prefix) to use with the map command.

--keyring    filename

Specifies a keyring file containing a secret for the specified user to use with the map command. If not specified, the default keyring locations will be searched.

--keyfile    filename

Specifies a file containing the secret key of `--id user` to use with the map command. This option is overridden by `--keyring` if the latter is also specified.

`--shared lock-tag`

Option for lock add that allows multiple clients to lock the same image if they use the same tag. The tag is an arbitrary string. This is useful for situations where an image must be open from more than one client at once, like during live migration of a virtual machine, or for use underneath a clustered file system.

`--format format`

Specifies output formatting (default: plain, json, xml)

`--pretty-format`

Make json or xml formatted output more human-readable.

`-o krbd-options``, --options krbd-options`

Specifies which options to use when mapping or unmapping an image via the rbd kernel driver. krbd-options is a comma-separated list of options (similar to mount(8) mount options). See kernel rbd (krbd) options section below for more details.

`--read-only`

Map the image read-only. Equivalent to -o ro.

`--image-feature feature-name`

Specifies which RBD format 2 feature should be enabled when creating an image. Multiple features can be enabled by repeating this option multiple times. The following features are supported:

- layering: layering support
- striping: striping v2 support
- exclusive-lock: exclusive locking support
- object-map: object map support (requires exclusive-lock)
- fast-diff: fast diff calculations (requires object-map)
- deep-flatten: snapshot flatten support
- journaling: journaled IO support (requires exclusive-lock)
- data-pool: erasure coded pool support

`--image-shared`

Specifies that the image will be used concurrently by multiple clients. This will disable features that are dependent upon exclusive ownership of the image.

`--whole-object`

Specifies that the diff should be limited to the extents of a full object instead of showing intra-object deltas. When the object map feature is enabled on an image, limiting the diff to the object extents will dramatically improve performance since the differences can be computed by examining the in-memory object map instead of querying RADOS for each object within the image.

```
--limit
```

Specifies the limit for the number of snapshots permitted.

## Commands

---

**bench** [-io-type <read | write | readwrite | rw> [-io-size size-in-B/K/M/G/T] [-io-threads num-ios-in-flight] [-io-total size-in-B/K/M/G/T] [-io-pattern seq | rand] [-rw-mix-read read proportion in readwrite] *image-spec*

Generate a series of IOs to the image and measure the IO throughput and latency. If no suffix is given, unit B is assumed for both -io-size and -io-total. Defaults are: -io-size 4096, -io-threads 16, -io-total 1G, -io-pattern seq, -rw-mix-read 50.

**children** *snap-spec*

List the clones of the image at the given snapshot. This checks every pool, and outputs the resulting poolname/imagename.

This requires image format 2.

**clone** [-object-size size-in-B/K/M] [-stripe-unit size-in-B/K/M -stripe-count num] [-image-feature feature-name] [-image-shared] *parent-snap-spec child-image-spec*

Will create a clone (copy-on-write child) of the parent snapshot. Object size will be identical to that of the parent image unless specified. Size will be the same as the parent snapshot. The -stripe-unit and -stripe-count arguments are optional, but must be used together.

The parent snapshot must be protected (see rbd snap protect). This requires image format 2.

**config global get** *config-entity key*

Get a global-level configuration override.

**config global list** [-format plain | json | xml] [-pretty-format] *config-entity*

List global-level configuration overrides.

**config global set** *config-entity key value*

Set a global-level configuration override.

**config global remove** *config-entity key*

Remove a global-level configuration override.

**config image get** *image-spec* *key*

Get an image-level configuration override.

**config image list** [-format plain | json | xml] [-pretty-format] *image-spec*

List image-level configuration overrides.

**config image set** *image-spec* *key* *value*

Set an image-level configuration override.

**config image remove** *image-spec* *key*

Remove an image-level configuration override.

**config pool get** *pool-name* *key*

Get a pool-level configuration override.

**config pool list** [-format plain | json | xml] [-pretty-format] *pool-name*

List pool-level configuration overrides.

**config pool set** *pool-name* *key* *value*

Set a pool-level configuration override.

**config pool remove** *pool-name* *key*

Remove a pool-level configuration override.

**cp** (*src-image-spec* | *src-snap-spec*) *dest-image-spec*

Copy the content of a *src-image* into the newly created *dest-image*. *dest-image* will have the same size, object size, and image format as *src-image*.

**create** (-s | -size *size-in-M/G/T*) [-image-format *format-id*] [-object-size *size-in-B/K/M*] [-stripe-unit *size-in-B/K/M* -stripe-count *num*] [-thick-provision] [-no-progress] [-image-feature *feature-name*]... [-image-shared] *image-spec*

Will create a new rbd image. You must also specify the size via -size. The -stripe-unit and -stripe-count arguments are optional, but must be used together. If the -thick-provision is enabled, it will fully allocate storage for the image at creation time. It will take a long time to do. Note: thick provisioning requires zeroing the contents of the entire image.

**deep cp** (*src-image-spec* | *src-snap-spec*) *dest-image-spec*

Deep copy the content of a *src-image* into the newly created *dest-image*. *Dest-image* will have the same size, object size, image format, and snapshots as *src-image*.

**device list** [-t | -device-type *device-type*] [-format plain | json | xml] -pretty-format

Show the rbd images that are mapped via the rbd kernel module (default) or other supported device.

**device map** [-t | -device-type *device-type*] [-read-only] [-exclusive] [-o | -options *device-options*] *image-spec* | *snap-spec*

Map the specified image to a block device via the rbd kernel module (default) or other supported device (*nbd* on Linux or *ggate* on FreeBSD).

The -options argument is a comma separated list of device type specific options (opt1,opt2=val,...).

**device unmap** [-t | -device-type *device-type*] [-o | -options *device-options*] *image-spec* | *snap-spec* | *device-path*

Unmap the block device that was mapped via the rbd kernel module (default) or other supported device.

The -options argument is a comma separated list of device type specific options (opt1,opt2=val,...).

**diff** [-from-snap *snap-name*] [-whole-object] *image-spec* | *snap-spec*

Dump a list of byte extents in the image that have changed since the specified start snapshot, or since the image was created. Each output line includes the starting offset (in bytes), the length of the region (in bytes), and either ‘zero’ or ‘data’ to indicate whether the region is known to be zeros or may contain other data.

**du** [-p | -pool *pool-name*] [*image-spec* | *snap-spec*] [-merge-snapshots]

Will calculate the provisioned and actual disk usage of all images and associated snapshots within the specified pool. It can also be used against individual images and snapshots.

If the RBD fast-diff feature is not enabled on images, this operation will require querying the OSDs for every potential object within the image.

The -merge-snapshots will merge snapshots used space into their parent images.

**export** [-export-format *format* (1 or 2)] (*image-spec* | *snap-spec*) [*dest-path*]

Export image to dest path (use - for stdout). The -export-format accepts ‘1’ or ‘2’ currently. Format 2 allow us to export not only the content of image, but also the snapshots and other properties, such as `image_order`, `features`.

**export-diff** [-from-snap *snap-name*] [-whole-object] (*image-spec* | *snap-spec*) *dest-path*

Export an incremental diff for an image to dest path (use - for stdout). If an initial

snapshot is specified, only changes since that snapshot are included; otherwise, any regions of the image that contain data are included. The end snapshot is specified using the standard `-snap` option or `@snap` syntax (see below). The image diff format includes metadata about image size changes, and the start and end snapshots. It efficiently represents discarded or ‘zero’ regions of the image.

**feature disable** *image-spec feature-name...*

Disable the specified feature on the specified image. Multiple features can be specified.

**feature enable** *image-spec feature-name...*

Enable the specified feature on the specified image. Multiple features can be specified.

**flatten** *image-spec*

If image is a clone, copy all shared blocks from the parent snapshot and make the child independent of the parent, severing the link between parent snap and child. The parent snapshot can be unprotected and deleted if it has no further dependent clones.

This requires image format 2.

**group create** *group-spec*

Create a group.

**group image add** *group-spec image-spec*

Add an image to a group.

**group image list** *group-spec*

List images in a group.

**group image remove** *group-spec image-spec*

Remove an image from a group.

**group ls** [-p | -pool *pool-name*]

List rbd groups.

**group rename** *src-group-spec dest-group-spec*

Rename a group. Note: rename across pools is not supported.

**group rm** *group-spec*

Delete a group.

**group snap create** *group-snap-spec*

Make a snapshot of a group.

**group snap list** *group-spec*

List snapshots of a group.

**group snap rm** *group-snap-spec*

Remove a snapshot from a group.

**group snap rename** *group-snap-spec* *snap-name*

Rename group's snapshot.

**group snap rollback** *group-snap-spec*

Rollback group to snapshot.

**image-meta get** *image-spec* *key*

Get metadata value with the key.

**image-meta list** *image-spec*

Show metadata held on the image. The first column is the key and the second column is the value.

**image-meta remove** *image-spec* *key*

Remove metadata key with the value.

**image-meta set** *image-spec* *key* *value*

Set metadata key with the value. They will displayed in image-meta list.

**import** [-export-format *format (1 or 2)*] [-image-format *format-id*] [-object-size *size-in-B/K/M*] [-stripe-unit *size-in-B/K/M* -stripe-count *num*] [-image-feature *feature-name*] ... [-image-shared] *src-path* [*image-spec*]

Create a new image and imports its data from path (use - for stdin). The import operation will try to create sparse rbd images if possible. For import from stdin, the sparsification unit is the data block size of the destination image (object size).

The -stripe-unit and -stripe-count arguments are optional, but must be used together.

The -export-format accepts '1' or '2' currently. Format 2 allow us to import not only the content of image, but also the snapshots and other properties, such as image\_order, features.

**import-diff** *src-path* *image-spec*

Import an incremental diff of an image and applies it to the current image. If the diff was generated relative to a start snapshot, we verify that snapshot already

exists before continuing. If there was an end snapshot we verify it does not already exist before applying the changes, and create the snapshot when we are done.

**info** *image-spec | snap-spec*

Will dump information (such as size and object size) about a specific rbd image. If image is a clone, information about its parent is also displayed. If a snapshot is specified, whether it is protected is shown as well.

**journal client disconnect** *journal-spec*

Flag image journal client as disconnected.

**journal export** [-verbose] [-no-error] *src-journal-spec path-name*

Export image journal to path (use - for stdout). It can be make a backup of the image journal especially before attempting dangerous operations.

Note that this command may not always work if the journal is badly corrupted.

**journal import** [-verbose] [-no-error] *path-name dest-journal-spec*

Import image journal from path (use - for stdin).

**journal info** *journal-spec*

Show information about image journal.

**journal inspect** [-verbose] *journal-spec*

Inspect and report image journal for structural errors.

**journal reset** *journal-spec*

Reset image journal.

**journal status** *journal-spec*

Show status of image journal.

**lock add** [-shared *lock-tag*] *image-spec lock-id*

Lock an image. The lock-id is an arbitrary name for the user's convenience. By default, this is an exclusive lock, meaning it will fail if the image is already locked. The -shared option changes this behavior. Note that locking does not affect any operation other than adding a lock. It does not protect an image from being deleted.

**lock ls** *image-spec*

Show locks held on the image. The first column is the locker to use with the lock remove command.

**lock rm** *image-spec lock-id locker*

Release a lock on an image. The lock id and locker are as output by lock ls.

**ls** [*-l* | *-long*] [*pool-name*]

Will list all rbd images listed in the rbd\_directory object. With *-l*, also show snapshots, and use longer-format output including size, parent (if clone), format, etc.

**merge-diff** *first-diff-path second-diff-path merged-diff-path*

Merge two continuous incremental diffs of an image into one single diff. The first diff's end snapshot must be equal with the second diff's start snapshot. The first diff could be - for stdin, and merged diff could be - for stdout, which enables multiple diff files to be merged using something like 'rbd merge-diff first second - | rbd merge-diff - third result'. Note this command currently only support the source incremental diff with stripe\_count == 1

**migration abort** *image-spec*

Cancel image migration. This step may be run after successful or failed migration prepare or migration execute steps and returns the image to its initial (before migration) state. All modifications to the destination image are lost.

**migration commit** *image-spec*

Commit image migration. This step is run after a successful migration prepare and migration execute steps and removes the source image data.

**migration execute** *image-spec*

Execute image migration. This step is run after a successful migration prepare step and copies image data to the destination.

**migration prepare** [*-order order*] [*-object-size object-size*] [*-image-feature image-feature*] [*-image-shared*] [*-stripe-unit stripe-unit*] [*-stripe-count stripe-count*] [*-data-pool data-pool*] [*-import-only*] [*-source-spec json*] [*-source-spec-path path*] *src-image-spec* [*dest-image-spec*]

Prepare image migration. This is the first step when migrating an image, i.e. changing the image location, format or other parameters that can't be changed dynamically. The destination can match the source, and in this case *dest-image-spec* can be omitted. After this step the source image is set as a parent of the destination image, and the image is accessible in copy-on-write mode by its destination spec.

An image can also be migrated from a read-only import source by adding the *-import-only* optional and providing a JSON-encoded *-source-spec* or a path to a JSON-encoded source-spec file using the *-source-spec-path* optionals.

**mirror image demote** *image-spec*

Demote a primary image to non-primary for RBD mirroring.

#### **mirror image disable [-force] *image-spec***

Disable RBD mirroring for an image. If the mirroring is configured in `image` mode for the image's pool, then it can be explicitly disabled mirroring for each image within the pool.

#### **mirror image enable *image-spec mode***

Enable RBD mirroring for an image. If the mirroring is configured in `image` mode for the image's pool, then it can be explicitly enabled mirroring for each image within the pool.

The mirror image mode can either be `journal` (default) or `snapshot`. The `journal` mode requires the RBD journaling feature.

#### **mirror image promote [-force] *image-spec***

Promote a non-primary image to primary for RBD mirroring.

#### **mirror image resync *image-spec***

Force resync to primary image for RBD mirroring.

#### **mirror image status *image-spec***

Show RBD mirroring status for an image.

#### **mirror pool demote [pool-name]**

Demote all primary images within a pool to non-primary. Every mirroring enabled image will be demoted in the pool.

#### **mirror pool disable [pool-name]**

Disable RBD mirroring by default within a pool. When mirroring is disabled on a pool in this way, mirroring will also be disabled on any images (within the pool) for which mirroring was enabled explicitly.

#### **mirror pool enable [pool-name] *mode***

Enable RBD mirroring by default within a pool. The mirroring mode can either be `pool` or `image`. If configured in `pool` mode, all images in the pool with the journaling feature enabled are mirrored. If configured in `image` mode, mirroring needs to be explicitly enabled (by `mirror image enable` command) on each image.

#### **mirror pool info [pool-name]**

Show information about the pool mirroring configuration. It includes mirroring mode, peer UUID, remote cluster name, and remote client name.

**mirror pool peer add [pool-name] remote-cluster-spec**

Add a mirroring peer to a pool. *remote-cluster-spec* is [*remote client name*@]*remote cluster name*.

The default for *remote client name* is “client.admin”.

This requires mirroring mode is enabled.

**mirror pool peer remove [pool-name] uuid**

Remove a mirroring peer from a pool. The peer *uuid* is available from **mirror pool info** command.

**mirror pool peer set [pool-name] uuid key value**

Update mirroring peer settings. The key can be either **client** or **cluster**, and the value is corresponding to remote client name or remote cluster name.

**mirror pool promote [-force] [pool-name]**

Promote all non-primary images within a pool to primary. Every mirroring enabled image will promoted in the pool.

**mirror pool status [-verbose] [pool-name]**

Show status for all mirrored images in the pool. With **-verbose**, also show additionally output status details for every mirroring image in the pool.

**mirror snapshot schedule add [-p | -pool pool] [-namespace namespace] [-image image] interval [start-time]**

Add mirror snapshot schedule.

**mirror snapshot schedule list [-R | -recursive] [-format format] [-pretty-format] [-p | -pool pool] [-namespace namespace] [-image image]**

List mirror snapshot schedule.

**mirror snapshot schedule remove [-p | -pool pool] [-namespace namespace] [-image image] interval [start-time]**

Remove mirror snapshot schedule.

**mirror snapshot schedule status [-p | -pool pool] [-format format] [-pretty-format] [-namespace namespace] [-image image]**

Show mirror snapshot schedule status.

**mv src-image-spec dest-image-spec**

Rename an image. Note: rename across pools is not supported.

```
namespace create pool-name/namespace-name
```

Create a new image namespace within the pool.

```
namespace list pool-name
```

List image namespaces defined within the pool.

```
namespace remove pool-name/namespace-name
```

Remove an empty image namespace from the pool.

```
object-map check image-spec | snap-spec
```

Verify the object map is correct.

```
object-map rebuild image-spec | snap-spec
```

Rebuild an invalid object map for the specified image. An image snapshot can be specified to rebuild an invalid object map for a snapshot.

```
pool init [pool-name] [-force]
```

Initialize pool for use by RBD. Newly created pools must initialized prior to use.

```
resize (-s | -size size-in-M/G/T) [-allow-shrink] image-spec
```

Resize rbd image. The size parameter also needs to be specified. The --allow-shrink option lets the size be reduced.

```
rm image-spec
```

Delete an rbd image (including all data blocks). If the image has snapshots, this fails and nothing is deleted.

```
snap create snap-spec
```

Create a new snapshot. Requires the snapshot name parameter specified.

```
snap limit clear image-spec
```

Remove any previously set limit on the number of snapshots allowed on an image.

```
snap limit set [-limit] limit image-spec
```

Set a limit for the number of snapshots allowed on an image.

```
snap ls image-spec
```

Dump the list of snapshots inside a specific image.

```
snap protect snap-spec
```

Protect a snapshot from deletion, so that clones can be made of it (see rbd clone).

Snapshots must be protected before clones are made; protection implies that there exist dependent cloned children that refer to this snapshot. rbd clone will fail on a nonprotected snapshot.

This requires image format 2.

**snap purge** *image-spec*

Remove all unprotected snapshots from an image.

**snap rename** *src-snap-spec* *dest-snap-spec*

Rename a snapshot. Note: rename across pools and images is not supported.

**snap rm** [*-force*] *snap-spec*

Remove the specified snapshot.

**snap rollback** *snap-spec*

Rollback image content to snapshot. This will iterate through the entire blocks array and update the data head content to the snapshotted version.

**snap unprotect** *snap-spec*

Unprotect a snapshot from deletion (undo snap protect). If cloned children remain, snap unprotect fails. (Note that clones may exist in different pools than the parent snapshot.)

This requires image format 2.

**sparsify** [*-sparse-size* *sparse-size*] *image-spec*

Reclaim space for zeroed image extents. The default sparse size is 4096 bytes and can be changed via *-sparse-size* option with the following restrictions: it should be power of two, not less than 4096, and not larger than image object size.

**status** *image-spec*

Show the status of the image, including which clients have it open.

**trash ls** [*pool-name*]

List all entries from trash.

**trash mv** *image-spec*

Move an image to the trash. Images, even ones actively in-use by clones, can be moved to the trash and deleted at a later time.

**trash purge** [*pool-name*]

Remove all expired images from trash.

**trash restore** *image-id*

Restore an image from trash.

**trash rm** *image-id*

Delete an image from trash. If image deferment time has not expired you can not removed it unless use force. But an actively in-use by clones or has snapshots can not be removed.

**trash purge schedule add** [-p | -pool *pool*] [-namespace *namespace*] *interval* [*start-time*]

Add trash purge schedule.

**trash purge schedule list** [-R | -recursive] [-format *format*] [-pretty-format] [-p | -pool *pool*] [-namespace *namespace*]

List trash purge schedule.

**trash purge schedule remove** [-p | -pool *pool*] [-namespace *namespace*] *interval* [*start-time*]

Remove trash purge schedule.

**trash purge schedule status** [-p | -pool *pool*] [-format *format*] [-pretty-format] [-namespace *namespace*]

Show trash purge schedule status.

**watch** *image-spec*

Watch events on image.

## Image, snap, group and journal specs

---

*image-spec* is [pool-name/[namespace-name/]]*image-name*

*snap-spec* is [pool-name/[namespace-name/]]*image-name@snap-name*

*group-spec* is [pool-name/[namespace-name/]]*group-name*

*group-snap-spec* is [pool-name/[namespace-name/]]*group-name@snap-name*

*journal-spec* is [pool-name/[namespace-name/]]*journal-name*

The default for *pool-name* is "rbd" and *namespace-name* is "". If an image name contains a slash character ('/'), *pool-name* is required.

The *journal-name* is *image-id*.

You may specify each name individually, using -pool, -namespace, -image, and -snap

options, but this is discouraged in favor of the above spec syntax.

## Striping

---

RBD images are striped over many objects, which are then stored by the Ceph distributed object store (RADOS). As a result, read and write requests for the image are distributed across many nodes in the cluster, generally preventing any single node from becoming a bottleneck when individual images get large or busy.

The striping is controlled by three parameters:

`object-size`

The size of objects we stripe over is a power of two. It will be rounded up the nearest power of two. The default object size is 4 MB, smallest is 4K and maximum is 32M.

`stripe_unit`

Each `[stripe_unit]` contiguous bytes are stored adjacently in the same object, before we move on to the next object.

`stripe_count`

After we write `[stripe_unit]` bytes to `[stripe_count]` objects, we loop back to the initial object and write another stripe, until the object reaches its maximum size. At that point, we move on to the next `[stripe_count]` objects.

By default, `[stripe_unit]` is the same as the object size and `[stripe_count]` is 1. Specifying a different `[stripe_unit]` and/or `[stripe_count]` is often referred to as using “fancy” striping and requires format 2.

## Kernel rbd (krbd) options

---

Most of these options are useful mainly for debugging and benchmarking. The default values are set in the kernel and may therefore depend on the version of the running kernel.

Per client instance rbd device map options:

- `fsid=aaaaaaaa-bbbb-cccc-dddd-eeeeeeeeeee` - FSID that should be assumed by the client.
- `ip=a.b.c.d[:p]` - IP and, optionally, port the client should use.
- `share` - Enable sharing of client instances with other mappings (default).
- `noshare` - Disable sharing of client instances with other mappings.
- `crc` - Enable CRC32C checksumming for data writes (default).

- `nocrc` - Disable CRC32C checksumming for data writes.
- `cephx_require_signatures` - Require cephx message signing (since 3.19, default).
- `nocephx_require_signatures` - Don't require cephx message signing (since 3.19).
- `tcp_nodelay` - Disable Nagle's algorithm on client sockets (since 4.0, default).
- `notcp_nodelay` - Enable Nagle's algorithm on client sockets (since 4.0).
- `cephx_sign_messages` - Enable message signing (since 4.4, default).
- `nocephx_sign_messages` - Disable message signing (since 4.4).
- `mount_timeout=x` - A timeout on various steps in rbd device map and rbd device unmap sequences (default is 60 seconds). In particular, since 4.2 this can be used to ensure that rbd device unmap eventually times out when there is no network connection to a cluster.
- `osdkeepalive=x` - OSD keepalive timeout (default is 5 seconds).
- `osd_idle_ttl=x` - OSD idle TTL (default is 60 seconds).

Per mapping (block device) rbd device map options:

- `rw` - Map the image read-write (default). Overridden by `-read-only`.
- `ro` - Map the image read-only. Equivalent to `-read-only`.
- `queue_depth=x` - queue depth (since 4.2, default is 128 requests).
- `lock_on_read` - Acquire exclusive lock on reads, in addition to writes and discards (since 4.9).
- `exclusive` - Disable automatic exclusive lock transitions (since 4.12). Equivalent to `-exclusive`.
- `lock_timeout=x` - A timeout on waiting for the acquisition of exclusive lock (since 4.17, default is 0 seconds, meaning no timeout).
- `notrim` - Turn off discard and write zeroes offload support to avoid deprovisioning a fully provisioned image (since 4.17). When enabled, discard requests will fail with `-EOPNOTSUPP`, write zeroes requests will fall back to manually zeroing.
- `abort_on_full` - Fail write requests with `-ENOSPC` when the cluster is full or the data pool reaches its quota (since 5.0). The default behaviour is to block until the full condition is cleared.
- `alloc_size` - Minimum allocation unit of the underlying OSD object store backend (since 5.1, default is 64K bytes). This is used to round off and drop discards that are too small. For bluestore, the recommended setting is

`bluestore_min_alloc_size` (typically 64K for hard disk drives and 16K for solid-state drives). For `filestore` with `filestore_punch_hole = false`, the recommended setting is image object size (typically 4M).

- `crush_location=x` - Specify the location of the client in terms of CRUSH hierarchy (since 5.8). This is a set of key-value pairs separated from each other by '|', with keys separated from values by ':'. Note that '|' may need to be quoted or escaped to avoid it being interpreted as a pipe by the shell. The key is the bucket type name (e.g. rack, datacenter or region with default bucket types) and the value is the bucket name. For example, to indicate that the client is local to rack "myrack", data center "mydc" and region "myregion":

```
1. crush_location=rack:myrack|datacenter:mydc|region:myregion
```

Each key-value pair stands on its own: "myrack" doesn't need to reside in "mydc", which in turn doesn't need to reside in "myregion". The location is not a path to the root of the hierarchy but rather a set of nodes that are matched independently, owing to the fact that bucket names are unique within a CRUSH map. "Multipath" locations are supported, so it is possible to indicate locality for multiple parallel hierarchies:

```
1. crush_location=rack:myrack1|rack:myrack2|datacenter:mydc
```

- `read_from_replica=no` - Disable replica reads, always pick the primary OSD (since 5.8, default).
- `read_from_replica=balance` - When issued a read on a replicated pool, pick a random OSD for serving it (since 5.8).

This mode is safe for general use only since Octopus (i.e. after "ceph osd require-osd-release octopus"). Otherwise it should be limited to read-only workloads such as images mapped read-only everywhere or snapshots.

- `read_from_replica=localize` - When issued a read on a replicated pool, pick the most local OSD for serving it (since 5.8). The locality metric is calculated against the location of the client given with `crush_location`; a match with the lowest-valued bucket type wins. For example, with default bucket types, an OSD in a matching rack is closer than an OSD in a matching data center, which in turn is closer than an OSD in a matching region.

This mode is safe for general use only since Octopus (i.e. after "ceph osd require-osd-release octopus"). Otherwise it should be limited to read-only workloads such as images mapped read-only everywhere or snapshots.

- `compression_hint=none` - Don't set compression hints (since 5.8, default).
- `compression_hint=compressible` - Hint to the underlying OSD object store backend that the data is compressible, enabling compression in passive mode (since 5.8).

- `compression_hint=incompressible` - Hint to the underlying OSD object store backend that the data is incompressible, disabling compression in aggressive mode (since 5.8).
- `udev` - Wait for udev device manager to finish executing all matching “add” rules and release the device before exiting (default). This option is not passed to the kernel.
- `noudev` - Don’t wait for udev device manager. When enabled, the device may not be fully usable immediately on exit.

rbd device unmap options:

- `force` - Force the unmapping of a block device that is open (since 4.9). The driver will wait for running requests to complete and then unmap; requests sent to the driver after initiating the unmap will be failed.
- `udev` - Wait for udev device manager to finish executing all matching “remove” rules and clean up after the device before exiting (default). This option is not passed to the kernel.
- `noudev` - Don’t wait for udev device manager.

## Examples

---

To create a new rbd image that is 100 GB:

```
1. rbd create mypool/myimage --size 102400
```

To use a non-default object size (8 MB):

```
1. rbd create mypool/myimage --size 102400 --object-size 8M
```

To delete an rbd image (be careful!):

```
1. rbd rm mypool/myimage
```

To create a new snapshot:

```
1. rbd snap create mypool/myimage@mysnap
```

To create a copy-on-write clone of a protected snapshot:

```
1. rbd clone mypool/myimage@mysnap otherpool/cloneimage
```

To see which clones of a snapshot exist:

```
1. rbd children mypool/myimage@mysnap
```

To delete a snapshot:

```
1. rbd snap rm mypool/myimage@mysnap
```

To map an image via the kernel with cephx enabled:

```
1. rbd device map mypool/myimage --id admin --keyfile secretfile
```

To map an image via the kernel with different cluster name other than default *ceph*:

```
1. rbd device map mypool/myimage --cluster cluster-name
```

To unmap an image:

```
1. rbd device unmap /dev/rbd0
```

To create an image and a clone from it:

```
1. rbd import --image-format 2 image mypool/parent
2. rbd snap create mypool/parent@snap
3. rbd snap protect mypool/parent@snap
4. rbd clone mypool/parent@snap otherpool/child
```

To create an image with a smaller stripe\_unit (to better distribute small writes in some workloads):

```
1. rbd create mypool/myimage --size 102400 --stripe-unit 65536B --stripe-count 16
```

To change an image from one image format to another, export it and then import it as the desired image format:

```
1. rbd export mypool/myimage@snap /tmp/img
2. rbd import --image-format 2 /tmp/img mypool/myimage2
```

To lock an image for exclusive use:

```
1. rbd lock add mypool/myimage mylockid
```

To release a lock:

```
1. rbd lock remove mypool/myimage mylockid client.2485
```

To list images from trash:

```
1. rbd trash ls mypool
```

To defer delete an image (use `--expires-at` to set expiration time, default is now):

```
1. rbd trash mv mypool/myimage --expires-at "tomorrow"
```

To delete an image from trash (be careful!):

```
1. rbd trash rm mypool/myimage-id
```

To force delete an image from trash (be careful!):

```
1. rbd trash rm mypool/myimage-id --force
```

To restore an image from trash:

```
1. rbd trash restore mypool/myimage-id
```

To restore an image from trash and rename it:

```
1. rbd trash restore mypool/myimage-id --image mynewimage
```

## Availability

**rbd** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[ceph\(8\)](#), [rados\(8\)](#)

# rbd-fuse – expose rbd images as files

## Synopsis

```
rbd-fuse [ -p pool ] [-c conffile] mountpoint [ fuse options ]
```

## Note

**rbd-fuse** is not recommended for any production or high performance workloads.

## Description

**rbd-fuse** is a FUSE (“Filesystem in USErspace”) client for RADOS block device (rbd) images. Given a pool containing rbd images, it will mount a userspace file system allowing access to those images as regular files at **mountpoint**.

The file system can be unmounted with:

```
1. fusermount -u mountpoint
```

or by sending **SIGINT** to the **rbd-fuse** process.

## Options

Any options not recognized by rbd-fuse will be passed on to libfuse.

```
-c ceph.conf
```

Use **ceph.conf** configuration file instead of the default **/etc/ceph/ceph.conf** to determine monitor addresses during startup.

```
-p pool
```

Use **pool** as the pool to search for rbd images. Default is **rbd**.

## Availability

**rbd-fuse** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[fusermount\(8\)](#), [rbd\(8\)](#)

# rbd-nbd – map rbd images to nbd device

## Synopsis

```
rbd-nbd [-c conf] [-read-only] [-device nbd device] [-nbds_max limit] [-max_part  
limit] [-exclusive] [-io-timeout seconds] [-reattach-timeout seconds] map image-spec |  
snap-spec
```

```
rbd-nbd unmap nbd device | image-spec | snap-spec
```

```
rbd-nbd list-mapped
```

```
rbd-nbd attach -device nbd device image-spec | snap-spec
```

```
rbd-nbd detach nbd device | image-spec | snap-spec
```

## Description

**rbd-nbd** is a client for RADOS block device (rbd) images like rbd kernel module. It will map a rbd image to a nbd (Network Block Device) device, allowing access it as regular local block device.

## Options

```
-c ceph.conf
```

Use `ceph.conf` configuration file instead of the default `/etc/ceph/ceph.conf` to determine monitor addresses during startup.

```
--read-only
```

Map read-only.

```
--nbds_max *limit*
```

Override the parameter of NBD kernel module when modprobe, used to limit the count of nbd device.

```
--max_part *limit*
```

Override for module param nbds\_max.

```
--exclusive
```

Forbid writes by other clients.

```
--io-timeout *seconds*
```

Override device timeout. Linux kernel will default to a 30 second request timeout.  
Allow the user to optionally specify an alternate timeout.

--reattach-timeout \*seconds\*

Specify timeout for the kernel to wait for a new rbd-nbd process is attached after the old process is detached. The default is 30 second.

## Image and snap specs

---

*image-spec* is [pool-name]/image-name

*snap-spec* is [pool-name]/image-name@snap-name

The default for *pool-name* is “rbd”. If an image name contains a slash character ('/'), *pool-name* is required.

## Availability

---

**rbd-nbd** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

---

[rbd\(8\)](#)

# rbd-ggate - map rbd images via FreeBSD GEOM Gate

## Synopsis

```
rbd-ggate [-read-only] [-exclusive] [-device ggate device] map image-spec | snap-spec  
rbd-ggate unmap ggate device  
rbd-ggate list
```

## Description

**rbd-ggate** is a client for RADOS block device (rbd) images. It will map a rbd image to a ggate (FreeBSD GEOM Gate class) device, allowing access it as regular local block device.

## Commands

### map

Spawn a process responsible for the creation of ggate device and forwarding I/O requests between the GEOM Gate kernel subsystem and RADOS.

### unmap

Destroy ggate device and terminate the process responsible for it.

### list

List mapped ggate devices.

## Options

--device \*ggate device\*

Specify ggate device path.

--read-only

Map read-only.

--exclusive

Forbid writes by other clients.

## Image and snap specs

---

*image-spec* is [pool-name]/image-name

*snap-spec* is [pool-name]/image-name@snap-name

The default for *pool-name* is “rbd”. If an image name contains a slash character ('/'), *pool-name* is required.

## Availability

---

**rbd-ggate** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

---

[rbd\(8\)](#) [ceph\(8\)](#)

# rbdmap – map RBD devices at boot time

## Synopsis

`rbdmap map`

`rbdmap unmap`

## Description

`rbdmap` is a shell script that automates `rbd map` and `rbd unmap` operations on one or more RBD (RADOS Block Device) images. While the script can be run manually by the system administrator at any time, the principal use case is automatic mapping/mounting of RBD images at boot time (and unmounting/unmapping at shutdown), as triggered by the init system (a systemd unit file, `rbdmap.service` is included with the ceph-common package for this purpose).

The script takes a single argument, which can be either “map” or “unmap”. In either case, the script parses a configuration file (defaults to `/etc/ceph/rbdmap`, but can be overridden via an environment variable `RBDMAPFILE`). Each line of the configuration file corresponds to an RBD image which is to be mapped, or unmapped.

The configuration file format is:

```
1. IMAGESPEC RBDOPTS
```

where `IMAGESPEC` should be specified as `POOLNAME/IMAGENAME` (the pool name, a forward slash, and the image name), or merely `IMAGENAME`, in which case the `POOLNAME` defaults to “rbd”. `RBDOPTS` is an optional list of parameters to be passed to the underlying `rbd map` command. These parameters and their values should be specified as a comma-separated string:

```
1. PARAM1=VAL1,PARAM2=VAL2,...,PARAMN=VALN
```

This will cause the script to issue an `rbd map` command like the following:

```
1. rbd map POOLNAME/IMAGENAME --PARAM1 VAL1 --PARAM2 VAL2
```

(See the `rbd` manpage for a full list of possible options.) For parameters and values which contain commas or equality signs, a simple apostrophe can be used to prevent replacing them.

When run as `rbdmap map`, the script parses the configuration file, and for each RBD

image specified attempts to first map the image (using the `rbd map` command) and, second, to mount the image.

When run as `rbdmap unmap`, images listed in the configuration file will be unmounted and unmapped.

`rbdmap unmap-all` attempts to unmount and subsequently unmap all currently mapped RBD images, regardless of whether or not they are listed in the configuration file.

If successful, the `rbd map` operation maps the image to a `/dev/rbdX` device, at which point a udev rule is triggered to create a friendly device name symlink, `/dev/rbd/POOLNAME/IMAGENAME`, pointing to the real mapped device.

In order for mounting/unmounting to succeed, the friendly device name must have a corresponding entry in `/etc/fstab`.

When writing `/etc/fstab` entries for RBD images, it's a good idea to specify the "noauto" (or "nofail") mount option. This prevents the init system from trying to mount the device too early - before the device in question even exists. (Since `rbdmap.service` executes a shell script, it is typically triggered quite late in the boot sequence.)

## Examples

---

Example `/etc/ceph/rbdmap` for three RBD images called "bar1", "bar2" and "bar3", which are in pool "foopool":

```
1. foopool/bar1    id=admin,keyring=/etc/ceph/ceph.client.admin.keyring
2. foopool/bar2    id=admin,keyring=/etc/ceph/ceph.client.admin.keyring
3. foopool/bar3    id=admin,keyring=/etc/ceph/ceph.client.admin.keyring,options='lock_on_read,queue_depth=1024'
```

Each line in the file contains two strings: the image spec and the options to be passed to `rbd map`. These two lines get transformed into the following commands:

```
1. rbd map foopool/bar1 --id admin --keyring /etc/ceph/ceph.client.admin.keyring
2. rbd map foopool/bar2 --id admin --keyring /etc/ceph/ceph.client.admin.keyring
   rbd map foopool/bar2 --id admin --keyring /etc/ceph/ceph.client.admin.keyring --options
3. lock_on_read,queue_depth=1024
```

If the images had XFS file systems on them, the corresponding `/etc/fstab` entries might look like this:

```
1. /dev/rbd/foopool/bar1 /mnt/bar1 xfs noauto 0 0
2. /dev/rbd/foopool/bar2 /mnt/bar2 xfs noauto 0 0
3. /dev/rbd/foopool/bar3 /mnt/bar3 xfs noauto 0 0
```

After creating the images and populating the `/etc/ceph/rbdmap` file, making the images get automatically mapped and mounted at boot is just a matter of enabling that unit:

```
1. systemctl enable rbdmap.service
```

# Options

---

None

## Availability

---

**rbdmap** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

---

[rbd\(8\)](#),

# ceph-rbdnamer – udev helper to name RBD devices

## Synopsis

```
ceph-rbdnamer num
```

## Description

**ceph-rbdnamer** prints the pool and image name for the given RBD devices to stdout. It is used by udev (using a rule like the one below) to set up a device symlink.

```
1. KERNEL=="rbd[0-9]*", PROGRAM="/usr/bin/ceph-rbdnamer %n", SYMLINK+="rbd/%c{1}/%c{2}"
```

## Availability

**ceph-rbdnamer** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[rbd\(8\)](#), [ceph\(8\)](#)

# rbd-replay-prep – prepare captured rados block device (RBD) workloads for replay

## Synopsis

```
rbd-replay-prep [ -window seconds ] [ -anonymize ] trace_dir replay_file
```

## Description

**rbd-replay-prep** processes raw rados block device (RBD) traces to prepare them for **rbd-replay**.

## Options

```
--window seconds
```

Requests further apart than ‘seconds’ seconds are assumed to be independent.

```
--anonymize
```

Anonymizes image and snap names.

```
--verbose
```

Print all processed events to console

## Examples

To prepare workload1-trace for replay:

```
1. rbd-replay-prep workload1-trace/ust/uid/1000/64-bit workload1
```

## Availability

**rbd-replay-prep** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[rbd-replay\(8\)](#), [rbd\(8\)](#)

# rbd-replay – replay rados block device (RBD) workloads

## Synopsis

```
rbd-replay [ options ] replay_file
```

## Description

**rbd-replay** is a utility for replaying rados block device (RBD) workloads.

## Options

```
-c ceph.conf``, --conf ceph.conf
```

Use `ceph.conf` configuration file instead of the default `/etc/ceph/ceph.conf` to determine monitor addresses during startup.

```
-p pool``, --pool pool
```

Interact with the given pool. Defaults to ‘`rbd`’.

```
--latency-multiplier
```

Multiplies inter-request latencies. Default: 1.

```
--read-only
```

Only replay non-destructive requests.

```
--map-image rule
```

Add a rule to map image names in the trace to image names in the replay cluster. A rule of `image1@snap1=image2@snap2` would map `snap1` of `image1` to `snap2` of `image2`.

```
--dump-perf-counters
```

**Experimental** Dump performance counters to standard out before an image is closed. Performance counters may be dumped multiple times if multiple images are closed, or if the same image is opened and closed multiple times. Performance counters and their meaning may change between versions.

## Examples

To replay workload1 as fast as possible:

```
1. rbd-replay --latency-multiplier=0 workload1
```

To replay workload1 but use test\_image instead of prod\_image:

```
1. rbd-replay --map-image=prod_image=test_image workload1
```

## Availability

**rbd-replay** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[rbd-replay-prep\(8\)](#), [rbd\(8\)](#)

# rbd-replay-many – replay a rados block device (RBD) workload on several clients

## Synopsis

```
rbd-replay-many [ options ] -original-image name host1 [ host2 [ ... ] ] -  
rbd_replay_args
```

## Description

**rbd-replay-many** is a utility for replaying a rados block device (RBD) workload on several clients. Although all clients use the same workload, they replay against separate images. This matches normal use of librbd, where each original client is a VM with its own image.

Configuration and replay files are not automatically copied to clients. Replay images must already exist.

## Options

```
--original-image name
```

Specifies the name (and snap) of the originally traced image. Necessary for correct name mapping.

```
--image-prefix prefix
```

Prefix of image names to replay against. Specifying `-image-prefix=foo` results in clients replaying against `foo-0`, `foo-1`, etc. Defaults to the original image name.

```
--exec program
```

Path to the rbd-replay executable.

```
--delay seconds
```

Delay between starting each client. Defaults to 0.

## Examples

Typical usage:

```
1. rbd-replay-many host-0 host-1 --original-image=image -- -c ceph.conf replay.bin
```

This results in the following commands being executed:

```
1. ssh host-0 'rbd-replay' --map-image 'image=image-0' -c ceph.conf replay.bin
2. ssh host-1 'rbd-replay' --map-image 'image=image-1' -c ceph.conf replay.bin
```

## Availability

**rbd-replay-many** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[rbd-replay\(8\)](#), [rbd\(8\)](#)

# Ceph Object Gateway

[Ceph Object Gateway](#) is an object storage interface built on top of [librados](#) to provide applications with a RESTful gateway to Ceph Storage Clusters. [Ceph Object Storage](#) supports two interfaces:

1. **S3-compatible:** Provides object storage functionality with an interface that is compatible with a large subset of the Amazon S3 RESTful API.
2. **Swift-compatible:** Provides object storage functionality with an interface that is compatible with a large subset of the OpenStack Swift API.

Ceph Object Storage uses the Ceph Object Gateway daemon ([radosgw](#)), which is an HTTP server for interacting with a Ceph Storage Cluster. Since it provides interfaces compatible with OpenStack Swift and Amazon S3, the Ceph Object Gateway has its own user management. Ceph Object Gateway can store data in the same Ceph Storage Cluster used to store data from Ceph File System clients or Ceph Block Device clients. The S3 and Swift APIs share a common namespace, so you may write data with one API and retrieve it with the other.



## Note

Ceph Object Storage does **NOT** use the Ceph Metadata Server.

- [HTTP Frontends](#)
- [Pool Placement and Storage Classes](#)
- [Multisite Configuration](#)
- [Multisite Sync Policy Configuration](#)
- [Configuring Pools](#)
- [Config Reference](#)
- [Admin Guide](#)
- [S3 API](#)
- [Data caching and CDN](#)
- [Swift API](#)
- [Admin Ops API](#)
- [Export over NFS](#)
- [OpenStack Keystone Integration](#)

- OpenStack Barbican Integration
- HashiCorp Vault Integration
- Open Policy Agent Integration
- Multi-tenancy
- Compression
- LDAP Authentication
- Server-Side Encryption
- Bucket Policy
- Dynamic bucket index resharding
- Multi factor authentication
- Sync Modules
- Bucket Notifications
- Data Layout in RADOS
- STS
- STS Lite
- Keycloak
- Role
- Orphan List and Associated Tooling
- OpenID Connect Provider
- Troubleshooting
- Manpage radosgw
- Manpage radosgw-admin
- QAT Acceleration for Encryption and Compression
- S3-select
- Lua Scripting

# HTTP Frontends

## Contents

- [HTTP Frontends](#)

- [Beast](#)

- [Options](#)

- [Civetweb](#)

- [Options](#)

- [Generic Options](#)

The Ceph Object Gateway supports two embedded HTTP frontend libraries that can be configured with `rgw_frontends`. See [Config Reference](#) for details about the syntax.

## Beast

New in version Mimic.

The `beast` frontend uses the Boost.Beast library for HTTP parsing and the Boost.Asio library for asynchronous network i/o.

## Options

`port` and `ssl_port`

Description

Sets the ipv4 & ipv6 listening port number. Can be specified multiple times as in

`port=80 port=8000`.

Type

Integer

Default

`80`

`endpoint` and `ssl_endpoint`

Description

Sets the listening address in the form `address[:port]`, where the address is an IPv4 address string in dotted decimal form, or an IPv6 address in hexadecimal notation

surrounded by square brackets. Specifying a IPv6 endpoint would listen to v6 only. The optional port defaults to 80 for `endpoint` and 443 for `ssl_endpoint`. Can be specified multiple times as in `endpoint=[::1] endpoint=192.168.0.100:8000`.

Type

Integer

Default

None

`ssl_certificate`

Description

Path to the SSL certificate file used for SSL-enabled endpoints. If path is prefixed with `config://`, the certificate will be pulled from the ceph monitor `config-key` database.

Type

String

Default

None

`ssl_private_key`

Description

Optional path to the private key file used for SSL-enabled endpoints. If one is not given, the `ssl_certificate` file is used as the private key. If path is prefixed with `config://`, the certificate will be pulled from the ceph monitor `config-key` database.

Type

String

Default

None

`tcp_nodelay`

Description

If set the socket option will disable Nagle's algorithm on the connection which means that packets will be sent as soon as possible instead of waiting for a full buffer or timeout to occur.

`1` Disable Nagel's algorithm for all sockets.

0 Keep the default: Nagel's algorithm enabled.

Type

Integer (0 or 1)

Default

0

max\_connection\_backlog

Description

Optional value to define the maximum size for the queue of connections waiting to be accepted. If not configured, the value from `boost::asio::socket_base::max_connections` will be used.

Type

Integer

Default

None

request\_timeout\_ms

Description

The amount of time in milliseconds that Beast will wait for more incoming data or outgoing data before giving up. Setting this value to 0 will disable timeout.

Type

Integer

Default

65000

## Civetweb

---

New in version Firefly.

The `civetweb` frontend uses the Civetweb HTTP library, which is a fork of Mongoose.

## Options

port

Description

Sets the listening port number. For SSL-enabled ports, add an `s` suffix like `443s`. To bind a specific IPv4 or IPv6 address, use the form `address:port`. Multiple endpoints can either be separated by `+` as in `127.0.0.1:8000+443s`, or by providing multiple options as in `port=8000 port=443s`.

Type

String

Default

`7480`

`num_threads`

Description

Sets the number of threads spawned by Civetweb to handle incoming HTTP connections. This effectively limits the number of concurrent connections that the frontend can service.

Type

Integer

Default

`rgw_thread_pool_size`

`request_timeout_ms`

Description

The amount of time in milliseconds that Civetweb will wait for more incoming data before giving up.

Type

Integer

Default

`30000`

`ssl_certificate`

Description

Path to the SSL certificate file used for SSL-enabled ports.

Type

String

Default

None

`access_log_file`

Description

Path to a file for access logs. Either full path, or relative to the current working directory. If absent (default), then accesses are not logged.

Type

String

Default

`EMPTY`

`error_log_file`

Description

Path to a file for error logs. Either full path, or relative to the current working directory. If absent (default), then errors are not logged.

Type

String

Default

`EMPTY`

The following is an example of the `/etc/ceph/ceph.conf` file with some of these options set:

```

1. [client.rgw.gateway-node1]
   rgw_frontends = civetweb request_timeout_ms=30000 error_log_file=/var/log/radosgw/civetweb.error.log
2. access_log_file=/var/log/radosgw/civetweb.access.log

```

A complete list of supported options can be found in the [Civetweb User Manual](#).

## Generic Options

Some frontend options are generic and supported by all frontends:

`prefix`

Description

A prefix string that is inserted into the URI of all requests. For example, a swift-only frontend could supply a uri prefix of `/swift`.

Type

String

Default

None

# Pool Placement and Storage Classes

## Contents

- Pool Placement and Storage Classes
  - Placement Targets
  - Storage Classes
  - Zonegroup/Zone Configuration
    - Adding a Placement Target
    - Adding a Storage Class
  - Customizing Placement
    - Default Placement
    - User Placement
    - S3 Bucket Placement
    - Swift Bucket Placement
  - Using Storage Classes

## Placement Targets

New in version Jewel.

Placement targets control which [Pools](#) are associated with a particular bucket. A bucket's placement target is selected on creation, and cannot be modified. The `radosgw-admin bucket stats` command will display its `placement_rule`.

The zonegroup configuration contains a list of placement targets with an initial target named `default-placement`. The zone configuration then maps each zonegroup placement target name onto its local storage. This zone placement information includes the `index_pool` name for the bucket index, the `data_extra_pool` name for metadata about incomplete multipart uploads, and a `data_pool` name for each storage class.

## Storage Classes

New in version Nautilus.

Storage classes are used to customize the placement of object data. S3 Bucket Lifecycle rules can automate the transition of objects between storage classes.

Storage classes are defined in terms of placement targets. Each zonegroup placement target lists its available storage classes with an initial class named `STANDARD`. The zone configuration is responsible for providing a `data_pool` pool name for each of the zonegroup's storage classes.

## Zonegroup/Zone Configuration

Placement configuration is performed with `radosgw-admin` commands on the zonegroups and zones.

The zonegroup placement configuration can be queried with:

```

1. $ radosgw-admin zonegroup get
2. {
3.     "id": "ab01123f-e0df-4f29-9d71-b44888d67cd5",
4.     "name": "default",
5.     "api_name": "default",
6.     ...
7.     "placement_targets": [
8.         {
9.             "name": "default-placement",
10.            "tags": [],
11.            "storage_classes": [
12.                "STANDARD"
13.            ]
14.        }
15.    ],
16.    "default_placement": "default-placement",
17.    ...
18. }
```

The zone placement configuration can be queried with:

```

1. $ radosgw-admin zone get
2. {
3.     "id": "557cdcee-3aae-4e9e-85c7-2f86f5eddb1f",
4.     "name": "default",
5.     "domain_root": "default.rgw.meta:root",
6.     ...
7.     "placement_pools": [
8.         {
9.             "key": "default-placement",
10.            "val": {
11.                "index_pool": "default.rgw.buckets.index",
12.                "storage_classes": {
13.                    "STANDARD": {
14.                        "data_pool": "default.rgw.buckets.data"
15.                    }
16.                },
17.                "data_extra_pool": "default.rgw.buckets.non-ec",
18.            }
19.        }
20.    ]
21. }
```

```

18.           "index_type": 0
19.       }
20.     }
21.   ],
22. ...
23. }
```

## Note

If you have not done any previous [Multisite Configuration](#), a `default` zone and zonegroup are created for you, and changes to the zone/zonegroup will not take effect until the Ceph Object Gateways are restarted. If you have created a realm for multisite, the zone/zonegroup changes will take effect once the changes are committed with `radosgw-admin period update --commit`.

## Adding a Placement Target

To create a new placement target named `temporary`, start by adding it to the zonegroup:

```

1. $ radosgw-admin zonegroup placement add \
2.   --rgw-zonegroup default \
3.   --placement-id temporary
```

Then provide the zone placement info for that target:

```

1. $ radosgw-admin zone placement add \
2.   --rgw-zone default \
3.   --placement-id temporary \
4.   --data-pool default.rgw.temporary.data \
5.   --index-pool default.rgw.temporary.index \
6.   --data-extra-pool default.rgw.temporary.non-ec
```

## Adding a Storage Class

To add a new storage class named `GLACIER` to the `default-placement` target, start by adding it to the zonegroup:

```

1. $ radosgw-admin zonegroup placement add \
2.   --rgw-zonegroup default \
3.   --placement-id default-placement \
4.   --storage-class GLACIER
```

Then provide the zone placement info for that storage class:

```

1. $ radosgw-admin zone placement add \
2.   --rgw-zone default \
3.   --placement-id default-placement \
4.   --storage-class GLACIER \
```

```

5.      --data-pool rgw.glacier.data \
6.      --compression lz4

```

## Customizing Placement

### Default Placement

By default, new buckets will use the zonegroup's `default_placement` target. This zonegroup setting can be changed with:

```

1. $ radosgw-admin zonegroup placement default \
2.   --rgw-zonegroup default \
3.   --placement-id new-placement

```

### User Placement

A Ceph Object Gateway user can override the zonegroup's default placement target by setting a non-empty `default_placement` field in the user info. Similarly, the `default_storage_class` can override the `STANDARD` storage class applied to objects by default.

```

1. $ radosgw-admin user info --uid testid
2. {
3.   ...
4.   "default_placement": "",
5.   "default_storage_class": "",
6.   "placement_tags": [],
7.   ...
8. }

```

If a zonegroup's placement target contains any `tags`, users will be unable to create buckets with that placement target unless their user info contains at least one matching tag in its `placement_tags` field. This can be useful to restrict access to certain types of storage.

The `radosgw-admin` command can modify these fields directly with:

```

1. $ radosgw-admin user modify \
2.   --uid <user-id> \
3.   --placement-id <default-placement-id> \
4.   --storage-class <default-storage-class> \
5.   --tags <tag1,tag2>

```

### S3 Bucket Placement

When creating a bucket with the S3 protocol, a placement target can be provided as part of the LocationConstraint to override the default placement targets from the user and zonegroup.

Normally, the LocationConstraint must match the zonegroup's `api_name` :

```
1. <LocationConstraint>default</LocationConstraint>
```

A custom placement target can be added to the `api_name` following a colon:

```
1. <LocationConstraint>default:new-placement</LocationConstraint>
```

## Swift Bucket Placement

When creating a bucket with the Swift protocol, a placement target can be provided in the HTTP header `X-Storage-Policy` :

```
1. X-Storage-Policy: new-placement
```

## Using Storage Classes

All placement targets have a `STANDARD` storage class which is applied to new objects by default. The user can override this default with its `default_storage_class` .

To create an object in a non-default storage class, provide that storage class name in an HTTP header with the request. The S3 protocol uses the `X-Amz-Storage-Class` header, while the Swift protocol uses the `X-Object-Storage-Class` header.

When using AWS S3 SDKs such as python boto3, it is important that the non-default storage class will be called as one of the AWS S3 allowed storage classes, or else the SDK will drop the request and raise an exception.

S3 Object Lifecycle Management can then be used to move object data between storage classes using `Transition` actions.

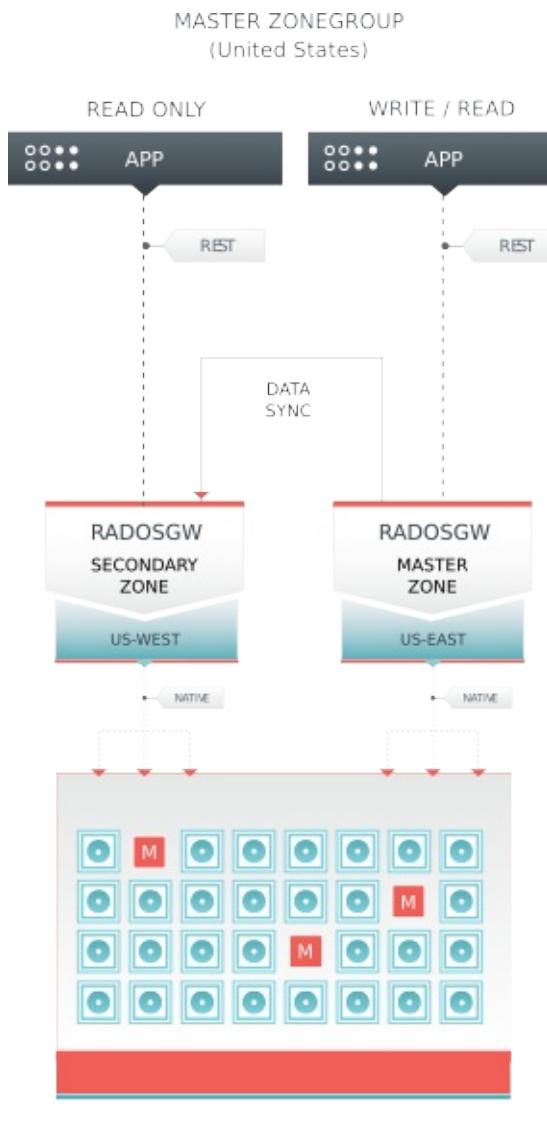
# Multi-Site

New in version Jewel.

A single zone configuration typically consists of one zone group containing one zone and one or more ceph-radosgw instances where you may load-balance gateway client requests between the instances. In a single zone configuration, typically multiple gateway instances point to a single Ceph storage cluster. However, Kraken supports several multi-site configuration options for the Ceph Object Gateway:

- **Multi-zone:** A more advanced configuration consists of one zone group and multiple zones, each zone with one or more ceph-radosgw instances. Each zone is backed by its own Ceph Storage Cluster. Multiple zones in a zone group provides disaster recovery for the zone group should one of the zones experience a significant failure. In Kraken, each zone is active and may receive write operations. In addition to disaster recovery, multiple active zones may also serve as a foundation for content delivery networks.
- **Multi-zone-group:** Formerly called ‘regions’, Ceph Object Gateway can also support multiple zone groups, each zone group with one or more zones. Objects stored to zones in one zone group within the same realm as another zone group will share a global object namespace, ensuring unique object IDs across zone groups and zones.
- **Multiple Realms:** In Kraken, the Ceph Object Gateway supports the notion of realms, which can be a single zone group or multiple zone groups and a globally unique namespace for the realm. Multiple realms provide the ability to support numerous configurations and namespaces.

Replicating object data between zones within a zone group looks something like this:



For additional details on setting up a cluster, see [Ceph Object Gateway for Production](#).

## Functional Changes from Infernalis

In Kraken, you can configure each Ceph Object Gateway to work in an active-active zone configuration, allowing for writes to non-master zones.

The multi-site configuration is stored within a container called a “realm.” The realm stores zone groups, zones, and a time “period” with multiple epochs for tracking changes to the configuration. In Kraken, the `ceph-radosgw` daemons handle the synchronization, eliminating the need for a separate synchronization agent. Additionally, the new approach to synchronization allows the Ceph Object Gateway to operate with an “active-active” configuration instead of “active-passive”.

## Requirements and Assumptions

A multi-site configuration requires at least two Ceph storage clusters, preferably

given a distinct cluster name. At least two Ceph object gateway instances, one for each Ceph storage cluster.

This guide assumes at least two Ceph storage clusters are in geographically separate locations; however, the configuration can work on the same site. This guide also assumes two Ceph object gateway servers named `rgw1` and `rgw2`.

### Important

Running a single Ceph storage cluster is NOT recommended unless you have low latency WAN connections.

A multi-site configuration requires a master zone group and a master zone. Additionally, each zone group requires a master zone. Zone groups may have one or more secondary or non-master zones.

In this guide, the `rgw1` host will serve as the master zone of the master zone group; and, the `rgw2` host will serve as the secondary zone of the master zone group.

See [Pools](#) for instructions on creating and tuning pools for Ceph Object Storage.

See [Sync Policy Config](#) for instructions on defining fine grained bucket sync policy rules.

## Configuring a Master Zone

All gateways in a multi-site configuration will retrieve their configuration from a `ceph-radosgw` daemon on a host within the master zone group and master zone. To configure your gateways in a multi-site configuration, choose a `ceph-radosgw` instance to configure the master zone group and master zone.

## Create a Realm

A realm contains the multi-site configuration of zone groups and zones and also serves to enforce a globally unique namespace within the realm.

Create a new realm for the multi-site configuration by opening a command line interface on a host identified to serve in the master zone group and zone. Then, execute the following:

```
1. # radosgw-admin realm create --rgw-realm={realm-name} [--default]
```

For example:

```
1. # radosgw-admin realm create --rgw-realm=movies --default
```

If the cluster will have a single realm, specify the `--default` flag. If `--default` is

specified, `radosgw-admin` will use this realm by default. If `--default` is not specified, adding zone-groups and zones requires specifying either the `--rgw-realm` flag or the `--realm-id` flag to identify the realm when adding zone groups and zones.

After creating the realm, `radosgw-admin` will echo back the realm configuration. For example:

```
1. {
2.   "id": "0956b174-fe14-4f97-8b50-bb7ec5e1cf62",
3.   "name": "movies",
4.   "current_period": "1950b710-3e63-4c41-a19e-46a715000980",
5.   "epoch": 1
6. }
```

#### Note

Ceph generates a unique ID for the realm, which allows the renaming of a realm if the need arises.

## Create a Master Zone Group

A realm must have at least one zone group, which will serve as the master zone group for the realm.

Create a new master zone group for the multi-site configuration by opening a command line interface on a host identified to serve in the master zone group and zone. Then, execute the following:

```
# radosgw-admin zonegroup create --rgw-zonegroup={name} --endpoints={url} [--rgw-realm={realm-name}] [--realm-
1. id={realm-id}] --master --default
```

For example:

```
# radosgw-admin zonegroup create --rgw-zonegroup=us --endpoints=http://rgw1:80 --rgw-realm=movies --master --
1. default
```

If the realm will only have a single zone group, specify the `--default` flag. If `--default` is specified, `radosgw-admin` will use this zone group by default when adding new zones. If `--default` is not specified, adding zones will require either the `--rgw-zonegroup` flag or the `--zonegroup-id` flag to identify the zone group when adding or modifying zones.

After creating the master zone group, `radosgw-admin` will echo back the zone group configuration. For example:

```
1. {
2.   "id": "f1a233f5-c354-4107-b36c-df66126475a6",
3.   "name": "us",
```

```

4.   "api_name": "us",
5.   "is_master": "true",
6.   "endpoints": [
7.     "http://rgw1:80"
8.   ],
9.   "hostnames": [],
10.  "hostnames_s3webzone": [],
11.  "master_zone": "",
12.  "zones": [],
13.  "placement_targets": [],
14.  "default_placement": "",
15.  "realm_id": "0956b174-fe14-4f97-8b50-bb7ec5e1cf62"
16. }

```

## Create a Master Zone

### Important

Zones must be created on a Ceph Object Gateway node that will be within the zone.

Create a new master zone for the multi-site configuration by opening a command line interface on a host identified to serve in the master zone group and zone. Then, execute the following:

```

1. # radosgw-admin zone create --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --master --default \
4.                               --endpoints={http://fqdn},{http://fqdn}

```

For example:

```

1. # radosgw-admin zone create --rgw-zonegroup=us --rgw-zone=us-east \
2.                               --master --default \
3.                               --endpoints={http://fqdn},{http://fqdn}

```

### Note

The `--access-key` and `--secret` aren't specified. These settings will be added to the zone once the user is created in the next section.

### Important

The following steps assume a multi-site configuration using newly installed systems that aren't storing data yet. DO NOT DELETE the `default` zone and its pools if you are already using it to store data, or the data will be deleted and unrecoverable.

## Delete Default Zone Group and Zone

Delete the `default` zone if it exists. Make sure to remove it from the default zone

group first.

```
1. # radosgw-admin zonegroup remove --rgw-zonegroup=default --rgw-zone=default
2. # radosgw-admin period update --commit
3. # radosgw-admin zone rm --rgw-zone=default
4. # radosgw-admin period update --commit
5. # radosgw-admin zonegroup delete --rgw-zonegroup=default
6. # radosgw-admin period update --commit
```

Finally, delete the `default` pools in your Ceph storage cluster if they exist.

#### Important

The following step assumes a multi-site configuration using newly installed systems that aren't currently storing data. DO NOT DELETE the `default` zone group if you are already using it to store data.

```
1. # ceph osd pool rm default.rgw.control default.rgw.control --yes-i-really-really-mean-it
2. # ceph osd pool rm default.rgw.data.root default.rgw.data.root --yes-i-really-really-mean-it
3. # ceph osd pool rm default.rgw.gc default.rgw.gc --yes-i-really-really-mean-it
4. # ceph osd pool rm default.rgw.log default.rgw.log --yes-i-really-really-mean-it
5. # ceph osd pool rm default.rgw.users.uid default.rgw.users.uid --yes-i-really-really-mean-it
```

## Create a System User

The `ceph-radosgw` daemons must authenticate before pulling realm and period information. In the master zone, create a system user to facilitate authentication between daemons.

```
1. # radosgw-admin user create --uid="{user-name}" --display-name="{Display Name}" --system
```

For example:

```
1. # radosgw-admin user create --uid="synchronization-user" --display-name="Synchronization User" --system
```

Make a note of the `access_key` and `secret_key`, as the secondary zones will require them to authenticate with the master zone.

Finally, add the system user to the master zone.

```
1. # radosgw-admin zone modify --rgw-zone=us-east --access-key={access-key} --secret={secret}
2. # radosgw-admin period update --commit
```

## Update the Period

After updating the master zone configuration, update the period.

```
1. # radosgw-admin period update --commit
```

#### Note

Updating the period changes the epoch, and ensures that other zones will receive the updated configuration.

## Update the Ceph Configuration File

Update the Ceph configuration file on master zone hosts by adding the `rgw_zone` configuration option and the name of the master zone to the instance entry.

```
1. [client.rgw.{instance-name}]
2. ...
3. rgw_zone={zone-name}
```

For example:

```
1. [client.rgw.rgw1]
2. host = rgw1
3. rgw frontends = "civetweb port=80"
4. rgw_zone=us-east
```

## Start the Gateway

On the object gateway host, start and enable the Ceph Object Gateway service:

```
1. # systemctl start ceph-radosgw@rgw.`hostname -s`
2. # systemctl enable ceph-radosgw@rgw.`hostname -s`
```

## Configure Secondary Zones

Zones within a zone group replicate all data to ensure that each zone has the same data. When creating the secondary zone, execute all of the following operations on a host identified to serve the secondary zone.

#### Note

To add a third zone, follow the same procedures as for adding the secondary zone. Use different zone name.

#### Important

You must execute metadata operations, such as user creation, on a host within the master zone. The master zone and the secondary zone can receive bucket operations, but the secondary zone redirects bucket operations to the master zone. If the master zone

is down, bucket operations will fail.

## Pull the Realm

Using the URL path, access key and secret of the master zone in the master zone group, pull the realm configuration to the host. To pull a non-default realm, specify the realm using the `--rgw-realm` or `--realm-id` configuration options.

```
1. # radosgw-admin realm pull --url={url-to-master-zone-gateway} --access-key={access-key} --secret={secret}
```

### Note

Pulling the realm also retrieves the remote's current period configuration, and makes it the current period on this host as well.

If this realm is the default realm or the only realm, make the realm the default realm.

```
1. # radosgw-admin realm default --rgw-realm={realm-name}
```

## Create a Secondary Zone

### Important

Zones must be created on a Ceph Object Gateway node that will be within the zone.

Create a secondary zone for the multi-site configuration by opening a command line interface on a host identified to serve the secondary zone. Specify the zone group ID, the new zone name and an endpoint for the zone. **DO NOT** use the `--master` or `--default` flags. In Kraken, all zones run in an active-active configuration by default; that is, a gateway client may write data to any zone and the zone will replicate the data to all other zones within the zone group. If the secondary zone should not accept write operations, specify the `--read-only` flag to create an active-passive configuration between the master zone and the secondary zone. Additionally, provide the `access_key` and `secret_key` of the generated system user stored in the master zone of the master zone group. Execute the following:

```
1. # radosgw-admin zone create --rgw-zonegroup={zone-group-name} \
2.           --rgw-zone={zone-name} --endpoints={url} \
3.           --access-key={system-key} --secret={secret} \
4.           --endpoints=http://{fqdn}:80 \
5.           [--read-only]
```

For example:

```
1. # radosgw-admin zone create --rgw-zonegroup=us --rgw-zone=us-west \
2.           --access-key={system-key} --secret={secret} \
```

```
3. --endpoints=http://rgw2:80
```

## Important

The following steps assume a multi-site configuration using newly installed systems that aren't storing data. **DO NOT DELETE** the `default` zone and its pools if you are already using it to store data, or the data will be lost and unrecoverable.

Delete the default zone if needed.

```
1. # radosgw-admin zone rm --rgw-zone=default
```

Finally, delete the default pools in your Ceph storage cluster if needed.

```
1. # ceph osd pool rm default.rgw.control default.rgw.control --yes-i-really-really-mean-it
2. # ceph osd pool rm default.rgw.data.root default.rgw.data.root --yes-i-really-really-mean-it
3. # ceph osd pool rm default.rgw.gc default.rgw.gc --yes-i-really-really-mean-it
4. # ceph osd pool rm default.rgw.log default.rgw.log --yes-i-really-really-mean-it
5. # ceph osd pool rm default.rgw.users.uid default.rgw.users.uid --yes-i-really-really-mean-it
```

## Update the Ceph Configuration File

Update the Ceph configuration file on the secondary zone hosts by adding the `rgw_zone` configuration option and the name of the secondary zone to the instance entry.

```
1. [client.rgw.{instance-name}]
2. ...
3. rgw_zone={zone-name}
```

For example:

```
1. [client.rgw.rgw2]
2. host = rgw2
3. rgw_frontends = "civetweb port=80"
4. rgw_zone=us-west
```

## Update the Period

After updating the master zone configuration, update the period.

```
1. # radosgw-admin period update --commit
```

## Note

Updating the period changes the epoch, and ensures that other zones will receive the updated configuration.

## Start the Gateway

On the object gateway host, start and enable the Ceph Object Gateway service:

```
1. # systemctl start ceph-radosgw@rgw.`hostname -s`  
2. # systemctl enable ceph-radosgw@rgw.`hostname -s`
```

## Check Synchronization Status

Once the secondary zone is up and running, check the synchronization status. Synchronization copies users and buckets created in the master zone to the secondary zone.

```
1. # radosgw-admin sync status
```

The output will provide the status of synchronization operations. For example:

```
1. realm f3239bc5-e1a8-4206-a81d-e1576480804d (earth)  
2.   zonegroup c50dbb7e-d9ce-47cc-a8bb-97d9b399d388 (us)  
3.     zone 4c453b70-4a16-4ce8-8185-1893b05d346e (us-west)  
4. metadata sync syncing  
5.       full sync: 0/64 shards  
6.       metadata is caught up with master  
7.       incremental sync: 64/64 shards  
8. data sync source: 1ee9da3e-114d-4ae3-a8a4-056e8a17f532 (us-east)  
9.       syncing  
10.      full sync: 0/128 shards  
11.      incremental sync: 128/128 shards  
12.      data is caught up with source
```

### Note

Secondary zones accept bucket operations; however, secondary zones redirect bucket operations to the master zone and then synchronize with the master zone to receive the result of the bucket operations. If the master zone is down, bucket operations executed on the secondary zone will fail, but object operations should succeed.

## Maintenance

### Checking the Sync Status

Information about the replication status of a zone can be queried with:

```
1. $ radosgw-admin sync status  
2.       realm b3bc1c37-9c44-4b89-a03b-04c269bea5da (earth)  
3.       zonegroup f54f9b22-b4b6-4a0e-9211-fa6ac1693f49 (us)
```

```

4.      zone adce11c9-b8ed-4a90-8bc5-3fc029ff0816 (us-2)
5.      metadata sync syncing
6.          full sync: 0/64 shards
7.          incremental sync: 64/64 shards
8.          metadata is behind on 1 shards
9.          oldest incremental change not applied: 2017-03-22 10:20:00.0.881361s
10.         data sync source: 341c2d81-4574-4d08-ab0f-5a2a7b168028 (us-1)
11.             syncing
12.             full sync: 0/128 shards
13.             incremental sync: 128/128 shards
14.             data is caught up with source
15.             source: 3b5d1a3f-3f27-4e4a-8f34-6072d4bb1275 (us-3)
16.                 syncing
17.                 full sync: 0/128 shards
18.                 incremental sync: 128/128 shards
19.                 data is caught up with source

```

## Changing the Metadata Master Zone

### Important

Care must be taken when changing which zone is the metadata master. If a zone has not finished syncing metadata from the current master zone, it will be unable to serve any remaining entries when promoted to master and those changes will be lost. For this reason, waiting for a zone's `radosgw-admin sync status` to catch up on metadata sync before promoting it to master is recommended.

Similarly, if changes to metadata are being processed by the current master zone while another zone is being promoted to master, those changes are likely to be lost. To avoid this, shutting down any `radosgw` instances on the previous master zone is recommended. After promoting another zone, its new period can be fetched with `radosgw-admin period pull` and the gateway(s) can be restarted.

To promote a zone (for example, zone `us-2` in zonegroup `us`) to metadata master, run the following commands on that zone:

```

1. $ radosgw-admin zone modify --rgw-zone=us-2 --master
2. $ radosgw-admin zonegroup modify --rgw-zonegroup=us --master
3. $ radosgw-admin period update --commit

```

This will generate a new period, and the `radosgw` instance(s) in zone `us-2` will send this period to other zones.

## Failover and Disaster Recovery

If the master zone should fail, failover to the secondary zone for disaster recovery.

1. Make the secondary zone the master and default zone. For example:

```
1. # radosgw-admin zone modify --rgw-zone={zone-name} --master --default
```

By default, Ceph Object Gateway will run in an active-active configuration. If the cluster was configured to run in an active-passive configuration, the secondary zone is a read-only zone. Remove the `--read-only` status to allow the zone to receive write operations. For example:

```
1. # radosgw-admin zone modify --rgw-zone={zone-name} --master --default \
2.           --read-only=false
```

2. Update the period to make the changes take effect.

```
1. # radosgw-admin period update --commit
```

3. Finally, restart the Ceph Object Gateway.

```
1. # systemctl restart ceph-radosgw@rgw.`hostname -s`
```

If the former master zone recovers, revert the operation.

1. From the recovered zone, pull the latest realm configuration from the current master zone.

```
1. # radosgw-admin realm pull --url={url-to-master-zone-gateway} \
2.           --access-key={access-key} --secret={secret}
```

2. Make the recovered zone the master and default zone.

```
1. # radosgw-admin zone modify --rgw-zone={zone-name} --master --default
```

3. Update the period to make the changes take effect.

```
1. # radosgw-admin period update --commit
```

4. Then, restart the Ceph Object Gateway in the recovered zone.

```
1. # systemctl restart ceph-radosgw@rgw.`hostname -s`
```

5. If the secondary zone needs to be a read-only configuration, update the secondary zone.

```
1. # radosgw-admin zone modify --rgw-zone={zone-name} --read-only
```

6. Update the period to make the changes take effect.

```
1. # radosgw-admin period update --commit
```

7. Finally, restart the Ceph Object Gateway in the secondary zone.

```
1. # systemctl restart ceph-radosgw@rgw.`hostname -s`
```

## Migrating a Single Site System to Multi-Site

To migrate from a single site system with a `default` zone group and zone to a multi site system, use the following steps:

1. Create a realm. Replace `<name>` with the realm name.

```
1. # radosgw-admin realm create --rgw-realm=<name> --default
```

2. Rename the default zone and zonegroup. Replace `<name>` with the zonegroup or zone name.

```
1. # radosgw-admin zonegroup rename --rgw-zonegroup default --zonegroup-new-name=<name>
2. # radosgw-admin zone rename --rgw-zone default --zone-new-name us-east-1 --rgw-zonegroup=<name>
```

3. Configure the master zonegroup. Replace `<name>` with the realm or zonegroup name. Replace `<fqdn>` with the fully qualified domain name(s) in the zonegroup.

```
# radosgw-admin zonegroup modify --rgw-realm=<name> --rgw-zonegroup=<name> --endpoints http://<fqdn>:80
1. --master --default
```

4. Configure the master zone. Replace `<name>` with the realm, zonegroup or zone name. Replace `<fqdn>` with the fully qualified domain name(s) in the zonegroup.

```
1. # radosgw-admin zone modify --rgw-realm=<name> --rgw-zonegroup=<name> \
2. --rgw-zone=<name> --endpoints http://<fqdn>:80 \
3. --access-key=<access-key> --secret=<secret-key> \
4. --master --default
```

5. Create a system user. Replace `<user-id>` with the username. Replace `<display-name>` with a display name. It may contain spaces.

```
1. # radosgw-admin user create --uid=<user-id> --display-name=<display-name> \
2. --access-key=<access-key> --secret=<secret-key> --system
```

6. Commit the updated configuration.

```
1. # radosgw-admin period update --commit
```

## 7. Finally, restart the Ceph Object Gateway.

```
1. # systemctl restart ceph-radosgw@rgw.`hostname -s`
```

After completing this procedure, proceed to [Configure a Secondary Zone](#) to create a secondary zone in the master zone group.

# Multi-Site Configuration Reference

The following sections provide additional details and command-line usage for realms, periods, zone groups and zones.

## Realms

A realm represents a globally unique namespace consisting of one or more zonegroups containing one or more zones, and zones containing buckets, which in turn contain objects. A realm enables the Ceph Object Gateway to support multiple namespaces and their configuration on the same hardware.

A realm contains the notion of periods. Each period represents the state of the zone group and zone configuration in time. Each time you make a change to a zonegroup or zone, update the period and commit it.

By default, the Ceph Object Gateway does not create a realm for backward compatibility with Infernalis and earlier releases. However, as a best practice, we recommend creating realms for new clusters.

## Create a Realm

To create a realm, execute `realm create` and specify the realm name. If the realm is the default, specify `--default`.

```
1. # radosgw-admin realm create --rgw-realm={realm-name} [--default]
```

For example:

```
1. # radosgw-admin realm create --rgw-realm=movies --default
```

By specifying `--default`, the realm will be called implicitly with each `radosgw-admin` call unless `--rgw-realm` and the realm name are explicitly provided.

## Make a Realm the Default

One realm in the list of realms should be the default realm. There may be only one default realm. If there is only one realm and it wasn't specified as the default realm when it was created, make it the default realm. Alternatively, to change which realm

is the default, execute:

```
1. # radosgw-admin realm default --rgw-realm=movies
```

## Note

When the realm is default, the command line assumes `--rgw-realm=<realm-name>` as an argument.

## Delete a Realm

To delete a realm, execute `realm delete` and specify the realm name.

```
1. # radosgw-admin realm delete --rgw-realm={realm-name}
```

For example:

```
1. # radosgw-admin realm delete --rgw-realm=movies
```

## Get a Realm

To get a realm, execute `realm get` and specify the realm name.

```
1. # radosgw-admin realm get --rgw-realm=<name>
```

For example:

```
1. # radosgw-admin realm get --rgw-realm=movies [> filename.json]
```

The CLI will echo a JSON object with the realm properties.

```
1. {
2.   "id": "0a68d52e-a19c-4e8e-b012-a8f831cb3ebc",
3.   "name": "movies",
4.   "current_period": "b0c5bbef-4337-4edd-8184-5aeab2ec413b",
5.   "epoch": 1
6. }
```

Use `>` and an output file name to output the JSON object to a file.

## Set a Realm

To set a realm, execute `realm set`, specify the realm name, and `--infile=` with an input file name.

```
1. # radosgw-admin realm set --rgw-realm=<name> --infile=<filename>
```

For example:

```
1. # radosgw-admin realm set --rgw-realm=movies --infile=filename.json
```

## List Realms

To list realms, execute `realm list`.

```
1. # radosgw-admin realm list
```

## List Realm Periods

To list realm periods, execute `realm list-periods`.

```
1. # radosgw-admin realm list-periods
```

## Pull a Realm

To pull a realm from the node containing the master zone group and master zone to a node containing a secondary zone group or zone, execute `realm pull` on the node that will receive the realm configuration.

```
1. # radosgw-admin realm pull --url={url-to-master-zone-gateway} --access-key={access-key} --secret={secret}
```

## Rename a Realm

A realm is not part of the period. Consequently, renaming the realm is only applied locally, and will not get pulled with `realm pull`. When renaming a realm with multiple zones, run the command on each zone. To rename a realm, execute the following:

```
1. # radosgw-admin realm rename --rgw-realm=<current-name> --realm-new-name=<new-realm-name>
```

### Note

DO NOT use `realm set` to change the `name` parameter. That changes the internal name only. Specifying `--rgw-realm` would still use the old realm name.

## Zone Groups

The Ceph Object Gateway supports multi-site deployments and a global namespace by using the notion of zone groups. Formerly called a region in Infernalis, a zone group defines the geographic location of one or more Ceph Object Gateway instances within one or more zones.

Configuring zone groups differs from typical configuration procedures, because not all

of the settings end up in a Ceph configuration file. You can list zone groups, get a zone group configuration, and set a zone group configuration.

## Create a Zone Group

Creating a zone group consists of specifying the zone group name. Creating a zone assumes it will live in the default realm unless `--rgw-realm=<realm-name>` is specified. If the zonegroup is the default zonegroup, specify the `--default` flag. If the zonegroup is the master zonegroup, specify the `--master` flag. For example:

```
1. # radosgw-admin zonegroup create --rgw-zonegroup=<name> [--rgw-realm=<name>][--master] [--default]
```

### Note

Use `zonegroup modify --rgw-zonegroup=<zonegroup-name>` to modify an existing zone group's settings.

## Make a Zone Group the Default

One zonegroup in the list of zonegroups should be the default zonegroup. There may be only one default zonegroup. If there is only one zonegroup and it wasn't specified as the default zonegroup when it was created, make it the default zonegroup.

Alternatively, to change which zonegroup is the default, execute:

```
1. # radosgw-admin zonegroup default --rgw-zonegroup=comedy
```

### Note

When the zonegroup is default, the command line assumes `--rgw-zonegroup=<zonegroup-name>` as an argument.

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## Add a Zone to a Zone Group

To add a zone to a zonegroup, execute the following:

```
1. # radosgw-admin zonegroup add --rgw-zonegroup=<name> --rgw-zone=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## Remove a Zone from a Zone Group

To remove a zone from a zonegroup, execute the following:

```
1. # radosgw-admin zonegroup remove --rgw-zonegroup=<name> --rgw-zone=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## Rename a Zone Group

To rename a zonegroup, execute the following:

```
1. # radosgw-admin zonegroup rename --rgw-zonegroup=<name> --zonegroup-new-name=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## Delete a Zone Group

To delete a zonegroup, execute the following:

```
1. # radosgw-admin zonegroup delete --rgw-zonegroup=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## List Zone Groups

A Ceph cluster contains a list of zone groups. To list the zone groups, execute:

```
1. # radosgw-admin zonegroup list
```

The `radosgw-admin` returns a JSON formatted list of zone groups.

```
1. {
2.     "default_info": "90b28698-e7c3-462c-a42d-4aa780d24eda",
3.     "zonegroups": [
4.         "us"
5.     ]
6. }
```

## Get a Zone Group Map

To list the details of each zone group, execute:

```
1. # radosgw-admin zonegroup-map get
```

## Note

If you receive a `failed to read zonegroup map` error, run `radosgw-admin zonegroup-map update` as `root` first.

## Get a Zone Group

To view the configuration of a zone group, execute:

```
1. radosgw-admin zonegroup get [--rgw-zonegroup=<zonegroup>]
```

The zone group configuration looks like this:

```
1. {
2.     "id": "90b28698-e7c3-462c-a42d-4aa780d24eda",
3.     "name": "us",
4.     "api_name": "us",
5.     "is_master": "true",
6.     "endpoints": [
7.         "http://rgw1:80"
8.     ],
9.     "hostnames": [],
10.    "hostnames_s3website": [],
11.    "master_zone": "9248cab2-afe7-43d8-a661-a40bf316665e",
12.    "zones": [
13.        {
14.            "id": "9248cab2-afe7-43d8-a661-a40bf316665e",
15.            "name": "us-east",
16.            "endpoints": [
17.                "http://rgw1"
18.            ],
19.            "log_meta": "true",
20.            "log_data": "true",
21.            "bucket_index_max_shards": 0,
22.            "read_only": "false"
23.        },
24.        {
25.            "id": "d1024e59-7d28-49d1-8222-af101965a939",
26.            "name": "us-west",
27.            "endpoints": [
28.                "http://rgw2:80"
29.            ],
30.            "log_meta": "false",
31.            "log_data": "true",
32.            "bucket_index_max_shards": 0,
33.            "read_only": "false"
34.        }
35.    ]
36. }
```

```

35.     ],
36.     "placement_targets": [
37.       {
38.         "name": "default-placement",
39.         "tags": []
40.       }
41.     ],
42.     "default_placement": "default-placement",
43.     "realm_id": "ae031368-8715-4e27-9a99-0c9468852cf"
44.   }

```

## Set a Zone Group

Defining a zone group consists of creating a JSON object, specifying at least the required settings:

1. `name` : The name of the zone group. Required.
2. `api_name` : The API name for the zone group. Optional.
3. `is_master` : Determines if the zone group is the master zone group. Required. **note:** You can only have one master zone group.
4. `endpoints` : A list of all the endpoints in the zone group. For example, you may use multiple domain names to refer to the same zone group. Remember to escape the forward slashes ( `\` ). You may also specify a port ( `fqdn:port` ) for each endpoint. Optional.
5. `hostnames` : A list of all the hostnames in the zone group. For example, you may use multiple domain names to refer to the same zone group. Optional. The `rgw dns name` setting will automatically be included in this list. You should restart the gateway daemon(s) after changing this setting.
6. `master_zone` : The master zone for the zone group. Optional. Uses the default zone if not specified. **note:** You can only have one master zone per zone group.
7. `zones` : A list of all zones within the zone group. Each zone has a name (required), a list of endpoints (optional), and whether or not the gateway will log metadata and data operations (false by default).
8. `placement_targets` : A list of placement targets (optional). Each placement target contains a name (required) for the placement target and a list of tags (optional) so that only users with the tag can use the placement target (i.e., the user's `placement_tags` field in the user info).
9. `default_placement` : The default placement target for the object index and object data. Set to `default-placement` by default. You may also set a per-user default placement in the user info for each user.

To set a zone group, create a JSON object consisting of the required fields, save the

object to a file (e.g., `zonegroup.json`) ; then, execute the following command:

```
1. # radosgw-admin zonegroup set --infile zonegroup.json
```

Where `zonegroup.json` is the JSON file you created.

### Important

The `default` zone group `is_master` setting is `true` by default. If you create a new zone group and want to make it the master zone group, you must either set the `default` zone group `is_master` setting to `false`, or delete the `default` zone group.

Finally, update the period:

```
1. # radosgw-admin period update --commit
```

## Set a Zone Group Map

Setting a zone group map consists of creating a JSON object consisting of one or more zone groups, and setting the `master_zonegroup` for the cluster. Each zone group in the zone group map consists of a key/value pair, where the `key` setting is equivalent to the `name` setting for an individual zone group configuration, and the `val` is a JSON object consisting of an individual zone group configuration.

You may only have one zone group with `is_master` equal to `true`, and it must be specified as the `master_zonegroup` at the end of the zone group map. The following JSON object is an example of a default zone group map.

```
1. {
2.   "zonegroups": [
3.     {
4.       "key": "90b28698-e7c3-462c-a42d-4aa780d24eda",
5.       "val": {
6.         "id": "90b28698-e7c3-462c-a42d-4aa780d24eda",
7.         "name": "us",
8.         "api_name": "us",
9.         "is_master": "true",
10.        "endpoints": [
11.          "http://rgw1:80"
12.        ],
13.        "hostnames": [],
14.        "hostnames_s3website": [],
15.        "master_zone": "9248cab2-afe7-43d8-a661-a40bf316665e",
16.        "zones": [
17.          {
18.            "id": "9248cab2-afe7-43d8-a661-a40bf316665e",
19.            "name": "us-east",
20.            "endpoints": [
21.              "http://rgw1"
22.            ],
23.          }
24.        ]
25.      }
26.    }
27.  ]
28.}
```

```

23.             "log_meta": "true",
24.             "log_data": "true",
25.             "bucket_index_max_shards": 0,
26.             "read_only": "false"
27.         },
28.         {
29.             "id": "d1024e59-7d28-49d1-8222-af101965a939",
30.             "name": "us-west",
31.             "endpoints": [
32.                 "http://rgw2:80"
33.             ],
34.             "log_meta": "false",
35.             "log_data": "true",
36.             "bucket_index_max_shards": 0,
37.             "read_only": "false"
38.         }
39.     ],
40.     "placement_targets": [
41.         {
42.             "name": "default-placement",
43.             "tags": []
44.         }
45.     ],
46.     "default_placement": "default-placement",
47.     "realm_id": "ae031368-8715-4e27-9a99-0c9468852cf"
48.   }
49. }
50. ],
51. "master_zonegroup": "90b28698-e7c3-462c-a42d-4aa780d24eda",
52. "bucket_quota": {
53.   "enabled": false,
54.   "max_size_kb": -1,
55.   "max_objects": -1
56. },
57. "user_quota": {
58.   "enabled": false,
59.   "max_size_kb": -1,
60.   "max_objects": -1
61. }
62. }

```

To set a zone group map, execute the following:

```
1. # radosgw-admin zonegroup-map set --infile zonegroupmap.json
```

Where `zonegroupmap.json` is the JSON file you created. Ensure that you have zones created for the ones specified in the zone group map. Finally, update the period.

```
1. # radosgw-admin period update --commit
```

## Zones

Ceph Object Gateway supports the notion of zones. A zone defines a logical group consisting of one or more Ceph Object Gateway instances.

Configuring zones differs from typical configuration procedures, because not all of the settings end up in a Ceph configuration file. You can list zones, get a zone configuration and set a zone configuration.

### Create a Zone

To create a zone, specify a zone name. If it is a master zone, specify the `--master` option. Only one zone in a zone group may be a master zone. To add the zone to a zonegroup, specify the `--rgw-zonegroup` option with the zonegroup name.

```
1. # radosgw-admin zone create --rgw-zone=<name> \
2.           [--zonegroup=<zonegroup-name>] \
3.           [--endpoints=<endpoint>[,<endpoint>] ] \
4.           [--master] [--default] \
5.           --access-key $SYSTEM_ACCESS_KEY --secret $SYSTEM_SECRET_KEY
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

### Delete a Zone

To delete zone, first remove it from the zonegroup.

```
1. # radosgw-admin zonegroup remove --zonegroup=<name> \
2.           --zone=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

Next, delete the zone. Execute the following:

```
1. # radosgw-admin zone rm --rgw-zone<name>
```

Finally, update the period:

```
1. # radosgw-admin period update --commit
```

### Important

Do not delete a zone without removing it from a zone group first. Otherwise, updating the period will fail.

If the pools for the deleted zone will not be used anywhere else, consider deleting the pools. Replace `<del-zone>` in the example below with the deleted zone's name.

### Important

Only delete the pools with prepended zone names. Deleting the root pool, such as, `.rgw.root` will remove all of the system's configuration.

### Important

Once the pools are deleted, all of the data within them are deleted in an unrecoverable manner. Only delete the pools if the pool contents are no longer needed.

```
1. # ceph osd pool rm <del-zone>.rgw.control <del-zone>.rgw.control --yes-i-really-really-mean-it
2. # ceph osd pool rm <del-zone>.rgw.data.root <del-zone>.rgw.data.root --yes-i-really-really-mean-it
3. # ceph osd pool rm <del-zone>.rgw.gc <del-zone>.rgw.gc --yes-i-really-really-mean-it
4. # ceph osd pool rm <del-zone>.rgw.log <del-zone>.rgw.log --yes-i-really-really-mean-it
5. # ceph osd pool rm <del-zone>.rgw.users.uid <del-zone>.rgw.users.uid --yes-i-really-really-mean-it
```

## Modify a Zone

To modify a zone, specify the zone name and the parameters you wish to modify.

```
1. # radosgw-admin zone modify [options]
```

Where `[options]` :

- `--access-key=<key>`
- `--secret/->secret-key=<key>`
- `--master`
- `--default`
- `--endpoints=<list>`

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## List Zones

As `root`, to list the zones in a cluster, execute:

```
1. # radosgw-admin zone list
```

## Get a Zone

As `root`, to get the configuration of a zone, execute:

```
1. # radosgw-admin zone get [--rgw-zone=<zone>]
```

The `default` zone looks like this:

```
1. { "domain_root": ".rgw",
2.   "control_pool": ".rgw.control",
3.   "gc_pool": ".rgw.gc",
4.   "log_pool": ".log",
5.   "intent_log_pool": ".intent-log",
6.   "usage_log_pool": ".usage",
7.   "user_keys_pool": ".users",
8.   "user_email_pool": ".users.email",
9.   "user_swift_pool": ".users.swift",
10.  "user_uid_pool": ".users.uid",
11.  "system_key": { "access_key": "", "secret_key": "" },
12.  "placement_pools": [
13.    { "key": "default-placement",
14.      "val": { "index_pool": ".rgw.buckets.index",
15.              "data_pool": ".rgw.buckets" }
16.    }
17.  ]
18. }
```

## Set a Zone

Configuring a zone involves specifying a series of Ceph Object Gateway pools. For consistency, we recommend using a pool prefix that is the same as the zone name. See [Pools](#) for details of configuring pools.

To set a zone, create a JSON object consisting of the pools, save the object to a file (e.g., `zone.json`); then, execute the following command, replacing `{zone-name}` with the name of the zone:

```
1. # radosgw-admin zone set --rgw-zone={zone-name} --infile zone.json
```

Where `zone.json` is the JSON file you created.

Then, as `root`, update the period:

```
1. # radosgw-admin period update --commit
```

## Rename a Zone

To rename a zone, specify the zone name and the new zone name.

```
1. # radosgw-admin zone rename --rgw-zone=<name> --zone-new-name=<name>
```

Then, update the period:

```
1. # radosgw-admin period update --commit
```

## Zone Group and Zone Settings

When configuring a default zone group and zone, the pool name includes the zone name. For example:

- `default.rgw.control`

To change the defaults, include the following settings in your Ceph configuration file under each `[client.radosgw.{instance-name}]` instance.

Name	Description	Type	Default
<code>rgw_zone</code>	The name of the zone for the gateway instance.	String	None
<code>rgw_zonegroup</code>	The name of the zone group for the gateway instance.	String	None
<code>rgw_zonegroup_root_pool</code>	The root pool for the zone group.	String	<code>.rgw.root</code>
<code>rgw_zone_root_pool</code>	The root pool for the zone.	String	<code>.rgw.root</code>
<code>rgw_default_zone_group_info_oid</code>	The OID for storing the default zone group. We do not recommend changing this setting.	String	<code>default.zonegroup</code>

# Multisite Sync Policy

New in version Octopus.

Multisite bucket-granularity sync policy provides fine grained control of data movement between buckets in different zones. It extends the zone sync mechanism. Previously buckets were being treated symmetrically, that is – each (data) zone holds a mirror of that bucket that should be the same as all the other zones. Whereas leveraging the bucket-granularity sync policy is possible for buckets to diverge, and a bucket can pull data from other buckets (ones that don't share its name or its ID) in different zone. The sync process was assuming therefore that the bucket sync source and the bucket sync destination were always referring to the same bucket, now that is not the case anymore.

The sync policy supersedes the old zonegroup coarse configuration (`sync_from*`). The sync policy can be configured at the zonegroup level (and if it is configured it replaces the old style config), but it can also be configured at the bucket level.

In the sync policy multiple groups that can contain lists of data-flow configurations can be defined, as well as lists of pipe configurations. The data-flow defines the flow of data between the different zones. It can define symmetrical data flow, in which multiple zones sync data from each other, and it can define directional data flow, in which the data moves in one way from one zone to another.

A pipe defines the actual buckets that can use these data flows, and the properties that are associated with it (for example: source object prefix).

A sync policy group can be in 3 states:

Value	Description
<code>enabled</code>	sync is allowed and enabled
<code>allowed</code>	sync is allowed
<code>forbidden</code>	sync (as defined by this group) is not allowed and can override other groups

A policy can be defined at the bucket level. A bucket level sync policy inherits the data flow of the zonegroup policy, and can only define a subset of what the zonegroup allows.

A wildcard zone, and a wildcard bucket parameter in the policy defines all relevant zones, or all relevant buckets. In the context of a bucket policy it means the current bucket instance. A disaster recovery configuration where entire zones are mirrored doesn't require configuring anything on the buckets. However, for a fine grained

bucket sync it would be better to configure the pipes to be synced by allowing (status=allowed) them at the zonegroup level (e.g., using wildcards), but only enable the specific sync at the bucket level (status=enabled). If needed, the policy at the bucket level can limit the data movement to specific relevant zones.

### Important

Any changes to the zonegroup policy needs to be applied on the zonegroup master zone, and require period update and commit. Changes to the bucket policy needs to be applied on the zonegroup master zone. The changes are dynamically handled by rgw.

## S3 Replication API

The S3 bucket replication api has also been implemented, and allows users to create replication rules between different buckets. Note though that while the AWS replication feature allows bucket replication within the same zone, rgw does not allow it at the moment. However, the rgw api also added a new 'Zone' array that allows users to select to what zones the specific bucket will be synced.

## Sync Policy Control Reference

### Get Sync Policy

To retrieve the current zonegroup sync policy, or a specific bucket policy:

```
1. # radosgw-admin sync policy get [--bucket=<bucket>]
```

### Create Sync Policy Group

To create a sync policy group:

```
1. # radosgw-admin sync group create [--bucket=<bucket>] \
2.           --group-id=<group-id> \
3.           --status=<enabled | allowed | forbidden> \
```

### Modify Sync Policy Group

To modify a sync policy group:

```
1. # radosgw-admin sync group modify [--bucket=<bucket>] \
2.           --group-id=<group-id> \
3.           --status=<enabled | allowed | forbidden> \
```

# Show Sync Policy Group

To show a sync policy group:

```
1. # radosgw-admin sync group get [--bucket=<bucket>]      \
2.                                --group-id=<group-id>
```

# Remove Sync Policy Group

To remove a sync policy group:

```
1. # radosgw-admin sync group remove [--bucket=<bucket>]      \
2.                                --group-id=<group-id>
```

# Create Sync Flow

- To create or update directional sync flow:

```
1. # radosgw-admin sync group flow create [--bucket=<bucket>]      \
2.                                --group-id=<group-id>      \
3.                                --flow-id=<flow-id>      \
4.                                --flow-type=directional      \
5.                                --source-zone=<source_zone>      \
6.                                --dest-zone=<dest_zone>
```

- To create or update symmetrical sync flow:

```
1. # radosgw-admin sync group flow create [--bucket=<bucket>]      \
2.                                --group-id=<group-id>      \
3.                                --flow-id=<flow-id>      \
4.                                --flow-type=symmetrical      \
5.                                --zones=<zones>
```

Where zones are a comma separated list of all the zones that need to add to the flow.

# Remove Sync Flow Zones

- To remove directional sync flow:

```
1. # radosgw-admin sync group flow remove [--bucket=<bucket>]      \
2.                                --group-id=<group-id>      \
3.                                --flow-id=<flow-id>      \
4.                                --flow-type=directional      \
5.                                --source-zone=<source_zone>      \
6.                                --dest-zone=<dest_zone>
```

- To remove specific zones from symmetrical sync flow:

```

1. # radosgw-admin sync group flow remove [--bucket=<bucket>]      \
2.                      --group-id=<group-id>          \
3.                      --flow-id=<flow-id>          \
4.                      --flow-type=symmetrical    \
5.                      --zones=<zones>

```

Where zones are a comma separated list of all zones to remove from the flow.

- To remove symmetrical sync flow:

```

1. # radosgw-admin sync group flow remove [--bucket=<bucket>]      \
2.                      --group-id=<group-id>          \
3.                      --flow-id=<flow-id>          \
4.                      --flow-type=symmetrical

```

## Create Sync Pipe

To create sync group pipe, or update its parameters:

```

1. # radosgw-admin sync group pipe create [--bucket=<bucket>]      \
2.                      --group-id=<group-id>          \
3.                      --pipe-id=<pipe-id>          \
4.                      --source-zones=<source_zones>    \
5.                      [--source-bucket=<source_buckets>]     \
6.                      [--source-bucket-id=<source_bucket_id>] \
7.                      --dest-zones=<dest_zones>        \
8.                      [--dest-bucket=<dest_buckets>]     \
9.                      [--dest-bucket-id=<dest_bucket_id>] \
10.                     [--prefix=<source_prefix>]       \
11.                     [--prefix-rm]                   \
12.                     [--tags-add=<tags>]           \
13.                     [--tags-rm=<tags>]           \
14.                     [--dest-owner=<owner>]         \
15.                     [--storage-class=<storage_class>] \
16.                     [--mode=<system | user>]        \
17.                     [--uid=<user_id>]

```

Zones are either a list of zones, or '\*' (wildcard). Wildcard zones mean any zone that matches the sync flow rules. Buckets are either a bucket name, or '\*' (wildcard). Wildcard bucket means the current bucket. Prefix can be defined to filter source objects. Tags are passed by a comma separated list of 'key=value'. Destination owner can be set to force a destination owner of the objects. If user mode is selected, only the destination bucket owner can be set. Destination storage class can also be configured. User id can be set for user mode, and will be the user under which the sync operation will be executed (for permissions validation).

# Remove Sync Pipe

To remove specific sync group pipe params, or the entire pipe:

```

1. # radosgw-admin sync group pipe remove [--bucket=<bucket>]           \
2.                                     --group-id=<group-id>                   \
3.                                     --pipe-id=<pipe-id>                     \
4.                                     [--source-zones=<source_zones>]      \
5.                                     [--source-bucket=<source_buckets>]    \
6.                                     [--source-bucket-id=<source_bucket_id>] \
7.                                     [--dest-zones=<dest_zones>]          \
8.                                     [--dest-bucket=<dest_buckets>]        \
9.                                     [--dest-bucket-id=<dest_bucket_id>]

```

# Sync Info

To get information about the expected sync sources and targets (as defined by the sync policy):

```

1. # radosgw-admin sync info [--bucket=<bucket>]           \
2.                                     [--effective-zone-name=<zone>]

```

Since a bucket can define a policy that defines data movement from it towards a different bucket at a different zone, when the policy is created we also generate a list of bucket dependencies that are used as hints when a sync of any particular bucket happens. The fact that a bucket references another bucket does not mean it actually syncs to/from it, as the data flow might not permit it.

## Examples

The system in these examples includes 3 zones: `us-east` (the master zone), `us-west`, `us-west-2`.

### Example 1: Two Zones, Complete Mirror

This is similar to older (pre `Octopus`) sync capabilities, but being done via the new sync policy engine. Note that changes to the zonegroup sync policy require a period update and commit.

```

1. [us-east] $ radosgw-admin sync group create --group-id=group1 --status=allowed
2. [us-east] $ radosgw-admin sync group flow create --group-id=group1 \
3.                                     --flow-id=flow-mirror --flow-type=symmetrical \
4.                                     --zones=us-east,us-west
5. [us-east] $ radosgw-admin sync group pipe create --group-id=group1 \
6.                                     --pipe-id=pipe1 --source-zones='*' \
7.                                     --source-bucket='*' --dest-zones='*' \

```

```

8.          --dest-bucket='*'
9. [us-east] $ radosgw-admin sync group modify --group-id=group1 --status=enabled
10. [us-east] $ radosgw-admin period update --commit
11.
12. $ radosgw-admin sync info --bucket=buck
13. {
14.     "sources": [
15.         {
16.             "id": "pipe1",
17.             "source": {
18.                 "zone": "us-west",
19.                 "bucket": "buck:115b12b3-....4409.1"
20.             },
21.             "dest": {
22.                 "zone": "us-east",
23.                 "bucket": "buck:115b12b3-....4409.1"
24.             },
25.             "params": {
26.             ...
27.             }
28.         }
29.     ],
30.     "dests": [
31.         {
32.             "id": "pipe1",
33.             "source": {
34.                 "zone": "us-east",
35.                 "bucket": "buck:115b12b3-....4409.1"
36.             },
37.             "dest": {
38.                 "zone": "us-west",
39.                 "bucket": "buck:115b12b3-....4409.1"
40.             },
41.             ...
42.         }
43.     ],
44.     ...
45.   }
46. }
```

Note that the “id” field in the output above reflects the pipe rule that generated that entry, a single rule can generate multiple sync entries as can be seen in the example.

```

1. [us-west] $ radosgw-admin sync info --bucket=buck
2. {
3.     "sources": [
4.         {
5.             "id": "pipe1",
6.             "source": {
7.                 "zone": "us-east",
8.                 "bucket": "buck:115b12b3-....4409.1"
```

```

9.          },
10.         "dest": {
11.             "zone": "us-west",
12.             "bucket": "buck:115b12b3-....4409.1"
13.         },
14.         ...
15.     }
16. ],
17. "dests": [
18.     {
19.         "id": "pipe1",
20.         "source": {
21.             "zone": "us-west",
22.             "bucket": "buck:115b12b3-....4409.1"
23.         },
24.         "dest": {
25.             "zone": "us-east",
26.             "bucket": "buck:115b12b3-....4409.1"
27.         },
28.         ...
29.     }
30. ],
31. ...
32. }
```

## Example 2: Directional, Entire Zone Backup

Also similar to older sync capabilities. In here we add a third zone, `us-west-2` that will be a replica of `us-west`, but data will not be replicated back from it.

```

1. [us-east] $ radosgw-admin sync group flow create --group-id=group1 \
2.                               --flow-id=us-west-backup --flow-type=directional \
3.                               --source-zone=us-west --dest-zone=us-west-2
4. [us-east] $ radosgw-admin period update --commit
```

Note that us-west has two dests:

```

1. [us-west] $ radosgw-admin sync info --bucket=buck
2. {
3.     "sources": [
4.         {
5.             "id": "pipe1",
6.             "source": {
7.                 "zone": "us-east",
8.                 "bucket": "buck:115b12b3-....4409.1"
9.             },
10.            "dest": {
11.                "zone": "us-west",
12.                "bucket": "buck:115b12b3-....4409.1"
13.            },
14.            ...
15.        }
16.    ],
17.    "dests": [
18.        {
19.            "id": "pipe2",
20.            "source": {
21.                "zone": "us-west-2",
22.                "bucket": "buck:115b12b3-....4409.1"
23.            },
24.            "dest": {
25.                "zone": "us-west-2",
26.                "bucket": "buck:115b12b3-....4409.1"
27.            }
28.        }
29.    ]
30. }
```

```

14.      ...
15.    }
16.  ],
17.  "dests": [
18.    {
19.      "id": "pipe1",
20.      "source": {
21.        "zone": "us-west",
22.        "bucket": "buck:115b12b3-....4409.1"
23.      },
24.      "dest": {
25.        "zone": "us-east",
26.        "bucket": "buck:115b12b3-....4409.1"
27.      },
28.      ...
29.    },
30.    {
31.      "id": "pipe1",
32.      "source": {
33.        "zone": "us-west",
34.        "bucket": "buck:115b12b3-....4409.1"
35.      },
36.      "dest": {
37.        "zone": "us-west-2",
38.        "bucket": "buck:115b12b3-....4409.1"
39.      },
40.      ...
41.    }
42.  ],
43.  ...
44. }
```

Whereas us-west-2 has only source and no destinations:

```

1. [us-west-2] $ radosgw-admin sync info --bucket=buck
2. {
3.   "sources": [
4.     {
5.       "id": "pipe1",
6.       "source": {
7.         "zone": "us-west",
8.         "bucket": "buck:115b12b3-....4409.1"
9.       },
10.      "dest": {
11.        "zone": "us-west-2",
12.        "bucket": "buck:115b12b3-....4409.1"
13.      },
14.      ...
15.    }
16.  ],
17.  "dests": [],
18.  ...
```

19. }

## Example 3: Mirror a Specific Bucket

Using the same group configuration, but this time switching it to `allowed` state, which means that sync is allowed but not enabled.

```
1. [us-east] $ radosgw-admin sync group modify --group-id=group1 --status=allowed
2. [us-east] $ radosgw-admin period update --commit
```

And we will create a bucket level policy rule for existing bucket `buck2`. Note that the bucket needs to exist before being able to set this policy, and admin commands that modify bucket policies need to run on the master zone, however, they do not require period update. There is no need to change the data flow, as it is inherited from the zonegroup policy. A bucket policy flow will only be a subset of the flow defined in the zonegroup policy. Same goes for pipes, although a bucket policy can enable pipes that are not enabled (albeit not forbidden) at the zonegroup policy.

```
1. [us-east] $ radosgw-admin sync group create --bucket=buck2 \
2.           --group-id=buck2-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync group pipe create --bucket=buck2 \
5.           --group-id=buck2-default --pipe-id=pipe1 \
6.           --source-zones='*' --dest-zones='*'
```

## Example 4: Limit Bucket Sync To Specific Zones

This will only sync `buck3` to `us-east` (from any zone that flow allows to sync into `us-east`).

```
1. [us-east] $ radosgw-admin sync group create --bucket=buck3 \
2.           --group-id=buck3-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync group pipe create --bucket=buck3 \
5.           --group-id=buck3-default --pipe-id=pipe1 \
6.           --source-zones='*' --dest-zones=us-east
```

## Example 5: Sync From a Different Bucket

Note that bucket sync only works (currently) across zones and not within the same zone.

Set `buck4` to pull data from `buck5`:

```
1. [us-east] $ radosgw-admin sync group create --bucket=buck4 '
```

```

2.                               --group-id=buck4-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync group pipe create --bucket=buck4 \
5.                               --group-id=buck4-default --pipe-id=pipe1 \
6.                               --source-zones='*' --source-bucket=buck5 \
7.                               --dest-zones='*'

```

can also limit it to specific zones, for example the following will only sync data originated in us-west:

```

1. [us-east] $ radosgw-admin sync group pipe modify --bucket=buck4 \
2.                               --group-id=buck4-default --pipe-id=pipe1 \
3.                               --source-zones=us-west --source-bucket=buck5 \
4.                               --dest-zones='*'

```

Checking the sync info for `buck5` on `us-west` is interesting:

```

1. [us-west] $ radosgw-admin sync info --bucket=buck5
2. {
3.     "sources": [],
4.     "dests": [],
5.     "hints": {
6.         "sources": [],
7.         "dests": [
8.             "buck4:115b12b3-....14433.2"
9.         ]
10.    },
11.    "resolved-hints-1": {
12.        "sources": [],
13.        "dests": [
14.            {
15.                "id": "pipe1",
16.                "source": {
17.                    "zone": "us-west",
18.                    "bucket": "buck5"
19.                },
20.                "dest": {
21.                    "zone": "us-east",
22.                    "bucket": "buck4:115b12b3-....14433.2"
23.                },
24.                ...
25.            },
26.            {
27.                "id": "pipe1",
28.                "source": {
29.                    "zone": "us-west",
30.                    "bucket": "buck5"
31.                },
32.                "dest": {
33.                    "zone": "us-west-2",
34.                    "bucket": "buck4:115b12b3-....14433.2"
35.                }
36.            }
37.        ]
38.    }
39. }

```

```

35.         },
36.         ...
37.     ],
38.   ],
39. },
40. "resolved-hints": {
41.   "sources": [],
42.   "dests": []
43. }
44. }
```

Note that there are resolved hints, which means that the bucket `buck5` found about `buck4` syncing from it indirectly, and not from its own policy (the policy for `buck5` itself is empty).

## Example 6: Sync To Different Bucket

The same mechanism can work for configuring data to be synced to (vs. synced from as in the previous example). Note that internally data is still pulled from the source at the destination zone:

Set `buck6` to “push” data to `buck5` :

```

1. [us-east] $ radosgw-admin sync group create --bucket=buck6 \
2.                               --group-id=buck6-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync group pipe create --bucket=buck6 \
5.                               --group-id=buck6-default --pipe-id=pipe1 \
6.                               --source-zones='*' --source-bucket='*' \
7.                               --dest-zones='*' --dest-bucket=buck5
```

A wildcard bucket name means the current bucket in the context of bucket sync policy.

Combined with the configuration in Example 5, we can now write data to `buck6` on `us-east`, data will sync to `buck5` on `us-west`, and from there it will be distributed to `buck4` on `us-east`, and on `us-west-2`.

## Example 7: Source Filters

Sync from `buck8` to `buck9`, but only objects that start with `foo/` :

```

1. [us-east] $ radosgw-admin sync group create --bucket=buck8 \
2.                               --group-id=buck8-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync group pipe create --bucket=buck8 \
5.                               --group-id=buck8-default --pipe-id=pipe-prefix \
6.                               --prefix=foo/ --source-zones='*' --dest-zones='*' \
7.                               --dest-bucket=buck9
```

Also sync from `buck8` to `buck9` any object that has the tags `color=blue` or `color:red` :

```

1. [us-east] $ radosgw-admin sync group pipe create --bucket=buck8 \
2.                               --group-id=buck8-default --pipe-id=pipe-tags \
3.                               --tags-add=color=blue,color=red --source-zones='*' \
4.                               --dest-zones='*' --dest-bucket=buck9

```

And we can check the expected sync in `us-east` (for example):

```

1. [us-east] $ radosgw-admin sync info --bucket=buck8
2. {
3.     "sources": [],
4.     "dests": [
5.         {
6.             "id": "pipe-prefix",
7.             "source": {
8.                 "zone": "us-east",
9.                 "bucket": "buck8:115b12b3-....14433.5"
10.            },
11.            "dest": {
12.                "zone": "us-west",
13.                "bucket": "buck9"
14.            },
15.            "params": {
16.                "source": {
17.                    "filter": {
18.                        "prefix": "foo/",
19.                        "tags": []
20.                    }
21.                },
22.                ...
23.            }
24.        },
25.        {
26.            "id": "pipe-tags",
27.            "source": {
28.                "zone": "us-east",
29.                "bucket": "buck8:115b12b3-....14433.5"
30.            },
31.            "dest": {
32.                "zone": "us-west",
33.                "bucket": "buck9"
34.            },
35.            "params": {
36.                "source": {
37.                    "filter": {
38.                        "tags": [
39.                            {
40.                                "key": "color",
41.                                "value": "blue"
42.                            },
43.                            {

```

```

44.           "key": "color",
45.           "value": "red"
46.       }
47.     ]
48.   }
49. },
50. ...
51.   }
52. }
53. ],
54. ...
55. }

```

Note that there aren't any sources, only two different destinations (one for each configuration). When the sync process happens it will select the relevant rule for each object it syncs.

Prefixes and tags can be combined, in which object will need to have both in order to be synced. The priority param can also be passed, and it can be used to determine when there are multiple different rules that are matched (and have the same source and destination), to determine which of the rules to be used.

## Example 8: Destination Params: Storage Class

Storage class of the destination objects can be configured:

```

1. [us-east] $ radosgw-admin sync group create --bucket=buck10 \
2.                               --group-id=buck10-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync pipe create --bucket=buck10 \
5.                               --group-id=buck10-default \
6.                               --pipe-id=pipe-storage-class \
7.                               --source-zones='*' --dest-zones=us-west-2 \
8.                               --storage-class=CHEAP_AND_SLOW

```

## Example 9: Destination Params: Destination Owner Translation

Set the destination objects owner as the destination bucket owner. This requires specifying the uid of the destination bucket:

```

1. [us-east] $ radosgw-admin sync group create --bucket=buck11 \
2.                               --group-id=buck11-default --status=enabled
3.
4. [us-east] $ radosgw-admin sync pipe create --bucket=buck11 \
5.                               --group-id=buck11-default --pipe-id=pipe-dest-owner \
6.                               --source-zones='*' --dest-zones='*' \
7.                               --dest-bucket=buck12 --dest-owner=joe

```

## Example 10: Destination Params: User Mode

User mode makes sure that the user has permissions to both read the objects, and write to the destination bucket. This requires that the uid of the user (which in its context the operation executes) is specified.

```
1. [us-east] $ radosgw-admin sync group pipe modify --bucket=buck11 \
2.                               --group-id=buck11-default --pipe-id=pipe-dest-owner \
3.                               --mode=user --uid=jenny
```

# Pools

The Ceph Object Gateway uses several pools for its various storage needs, which are listed in the Zone object (see `radosgw-admin zone get`). A single zone named `default` is created automatically with pool names starting with `default.rgw.`, but a [Multisite Configuration](#) will have multiple zones.

## Tuning

When `radosgw` first tries to operate on a zone pool that does not exist, it will create that pool with the default values from `osd pool default pg num` and `osd pool default pgp num`. These defaults are sufficient for some pools, but others (especially those listed in `placement_pools` for the bucket index and data) will require additional tuning. We recommend using the [Ceph Placement Group's per Pool Calculator](#) to calculate a suitable number of placement groups for these pools. See [Pools](#) for details on pool creation.

## Pool Namespaces

New in version Luminous.

Pool names particular to a zone follow the naming convention `{zone-name}.pool-name`. For example, a zone named `us-east` will have the following pools:

- `.rgw.root`
- `us-east.rgw.control`
- `us-east.rgw.meta`
- `us-east.rgw.log`
- `us-east.rgw.buckets.index`
- `us-east.rgw.buckets.data`

The zone definitions list several more pools than that, but many of those are consolidated through the use of rados namespaces. For example, all of the following pool entries use namespaces of the `us-east.rgw.meta` pool:

```

1. "user_keys_pool": "us-east.rgw.meta:users.keys",
2. "user_email_pool": "us-east.rgw.meta:users.email",
3. "user_swift_pool": "us-east.rgw.meta:users.swift",
4. "user_uid_pool": "us-east.rgw.meta:users.uid",

```

# Ceph Object Gateway Config Reference

The following settings may be added to the Ceph configuration file (i.e., usually `ceph.conf`) under the `[client.radosgw.{instance-name}]` section. The settings may contain default values. If you do not specify each setting in the Ceph configuration file, the default value will be set automatically.

Configuration variables set under the `[client.radosgw.{instance-name}]` section will not apply to rgw or radosgw-admin commands without an instance-name specified in the command. Thus variables meant to be applied to all RGW instances or all radosgw-admin commands can be put into the `[global]` or the `[client]` section to avoid specifying instance-name.

## `rgw frontends`

### Description

Configures the HTTP frontend(s). The configuration for multiple frontends can be provided in a comma-delimited list. Each frontend configuration may include a list of options separated by spaces, where each option is in the form “key=value” or “key”. See [HTTP Frontends](#) for more on supported options.

### Type

String

### Default

```
beast port=7480
```

```
rgw data
```

### Description

Sets the location of the data files for Ceph Object Gateway.

### Type

String

### Default

```
/var/lib/ceph/radosgw/$cluster-$id
```

```
rgw enable apis
```

### Description

Enables the specified APIs.

### Note

Enabling the `s3` API is a requirement for any radosgw instance that is meant to participate in a [multi-site](#) configuration.

Type

String

Default

`s3, swift, swift_auth, admin` All APIs.

`rgw cache enabled`

Description

Whether the Ceph Object Gateway cache is enabled.

Type

Boolean

Default

`true`

`rgw cache lru size`

Description

The number of entries in the Ceph Object Gateway cache.

Type

Integer

Default

`10000`

`rgw socket path`

Description

The socket path for the domain socket. `FastCgiExternalServer` uses this socket. If you do not specify a socket path, Ceph Object Gateway will not run as an external server. The path you specify here must be the same as the path specified in the `rgw.conf` file.

Type

String

Default

N/A

`rgw fcgi socket backlog`

## Description

The socket backlog for fcgi.

## Type

Integer

## Default

1024

rgw host

## Description

The host for the Ceph Object Gateway instance. Can be an IP address or a hostname.

## Type

String

## Default

0.0.0.0

rgw port

## Description

Port the instance listens for requests. If not specified, Ceph Object Gateway runs external FastCGI.

## Type

String

## Default

None

rgw dns name

## Description

The DNS name of the served domain. See also the `hostnames` setting within regions.

## Type

String

## Default

None

rgw script uri

## Description

The alternative value for the `SCRIPT_URI` if not set in the request.

### Type

String

### Default

None

```
rgw request uri
```

## Description

The alternative value for the `REQUEST_URI` if not set in the request.

### Type

String

### Default

None

```
rgw print continue
```

## Description

Enable `100-continue` if it is operational.

### Type

Boolean

### Default

```
true
```

```
rgw remote addr param
```

## Description

The remote address parameter. For example, the HTTP field containing the remote address, or the `X-Forwarded-For` address if a reverse proxy is operational.

### Type

String

### Default

```
REMOTE_ADDR
```

```
rgw op thread timeout
```

## Description

The timeout in seconds for open threads.

## Type

Integer

## Default

600

```
rgw op thread suicide timeout
```

## Description

The time `timeout` in seconds before a Ceph Object Gateway process dies. Disabled if set to `0`.

## Type

Integer

## Default

```
0
```

```
rgw thread pool size
```

## Description

The size of the thread pool.

## Type

Integer

## Default

100 threads.

```
rgw num control oids
```

## Description

The number of notification objects used for cache synchronization between different `rgw` instances.

## Type

Integer

## Default

```
8
```

```
rgw init timeout
```

#### Description

The number of seconds before Ceph Object Gateway gives up on initialization.

#### Type

Integer

#### Default

```
30
```

```
rgw mime types file
```

#### Description

The path and location of the MIME types. Used for Swift auto-detection of object types.

#### Type

String

#### Default

```
/etc/mime.types
```

```
rgw s3 success create obj status
```

#### Description

The alternate success status response for `create-obj`.

#### Type

Integer

#### Default

```
0
```

```
rgw resolve cname
```

#### Description

Whether `rgw` should use DNS CNAME record of the request hostname field (if hostname is not equal to `rgw dns name`).

#### Type

Boolean

#### Default

`false``rgw obj stripe size`

### Description

The size of an object stripe for Ceph Object Gateway objects. See [Architecture](#) for details on striping.

### Type

Integer

### Default

`4 << 20``rgw extended http attrs`

### Description

Add new set of attributes that could be set on an entity (user, bucket or object). These extra attributes can be set through HTTP header fields when putting the entity or modifying it using POST method. If set, these attributes will return as HTTP fields when doing GET/HEAD on the entity.

### Type

String

### Default

None

### Example

“content\_foo, content\_bar, x-foo-bar”

`rgw exit timeout secs`

### Description

Number of seconds to wait for a process before exiting unconditionally.

### Type

Integer

### Default

`120``rgw get obj window size`

### Description

The window size in bytes for a single object request.

Type

Integer

Default

16 << 20

rgw get obj max req size

Description

The maximum request size of a single get operation sent to the Ceph Storage Cluster.

Type

Integer

Default

4 << 20

rgw relaxed s3 bucket names

Description

Enables relaxed S3 bucket names rules for US region buckets.

Type

Boolean

Default

false

rgw list buckets max chunk

Description

The maximum number of buckets to retrieve in a single operation when listing user buckets.

Type

Integer

Default

1000

rgw override bucket index max shards

Description

Represents the number of shards for the bucket index object, a value of zero indicates there is no sharding. It is not recommended to set a value too large (e.g. thousand) as it increases the cost for bucket listing. This variable should be set in the client or global sections so that it is automatically applied to radosgw-admin commands.

Type

Integer

Default

0

```
rgw curl wait timeout ms
```

Description

The timeout in milliseconds for certain `curl` calls.

Type

Integer

Default

1000

```
rgw copy obj progress
```

Description

Enables output of object progress during long copy operations.

Type

Boolean

Default

true

```
rgw copy obj progress every bytes
```

Description

The minimum bytes between copy progress output.

Type

Integer

Default

1024 \* 1024

```
rgw admin entry
```

## Description

The entry point for an admin request URL.

## Type

String

## Default

admin

rgw content length compat

## Description

Enable compatibility handling of FCGI requests with both CONTENT\_LENGTH AND HTTP\_CONTENT\_LENGTH set.

## Type

Boolean

## Default

false

rgw bucket quota ttl

## Description

The amount of time in seconds cached quota information is trusted. After this timeout, the quota information will be re-fetched from the cluster.

## Type

Integer

## Default

600

rgw user quota bucket sync interval

## Description

The amount of time in seconds bucket quota information is accumulated before syncing to the cluster. During this time, other RGW instances will not see the changes in bucket quota stats from operations on this instance.

## Type

Integer

## Default

`180``rgw user quota sync interval`

#### Description

The amount of time in seconds user quota information is accumulated before syncing to the cluster. During this time, other RGW instances will not see the changes in user quota stats from operations on this instance.

#### Type

Integer

#### Default

`180``rgw bucket default quota max objects`

#### Description

Default max number of objects per bucket. Set on new users, if no other quota is specified. Has no effect on existing users. This variable should be set in the client or global sections so that it is automatically applied to radosgw-admin commands.

#### Type

Integer

#### Default

`-1``rgw bucket default quota max size`

#### Description

Default max capacity per bucket, in bytes. Set on new users, if no other quota is specified. Has no effect on existing users.

#### Type

Integer

#### Default

`-1``rgw user default quota max objects`

#### Description

Default max number of objects for a user. This includes all objects in all buckets owned by the user. Set on new users, if no other quota is specified. Has no effect on

existing users.

Type

Integer

Default

-1

```
rgw user default quota max size
```

Description

The value for user max size quota in bytes set on new users, if no other quota is specified. Has no effect on existing users.

Type

Integer

Default

-1

```
rgw verify ssl
```

Description

Verify SSL certificates while making requests.

Type

Boolean

Default

true

## Lifecycle Settings

Bucket Lifecycle configuration can be used to manage your objects so they are stored effectively throughout their lifetime. In past releases Lifecycle processing was rate-limited by single threaded processing. With the Nautilus release this has been addressed and the Ceph Object Gateway now allows for parallel thread processing of bucket lifecycles across additional Ceph Object Gateway instances and replaces the in-order index shard enumeration with a random ordered sequence.

There are two options in particular to look at when looking to increase the aggressiveness of lifecycle processing:

```
rgw lc max worker
```

## Description

This option specifies the number of lifecycle worker threads to run in parallel, thereby processing bucket and index shards simultaneously.

## Type

Integer

## Default

3

`rgw lc max wp worker`

## Description

This option specifies the number of threads in each lifecycle workers work pool. This option can help accelerate processing each bucket.

These values can be tuned based upon your specific workload to further increase the aggressiveness of lifecycle processing. For a workload with a larger number of buckets (thousands) you would look at increasing the `rgw lc max worker` value from the default value of 3 whereas a workload with a smaller number of buckets but higher number of objects (hundreds of thousands) per bucket you would look at tuning `rgw lc max wp worker` from the default value of 3.

## NOTE

When looking to tune either of these specific values please validate the current cluster performance and Ceph Object Gateway utilization before increasing.

# Garbage Collection Settings

The Ceph Object Gateway allocates storage for new objects immediately.

The Ceph Object Gateway purges the storage space used for deleted and overwritten objects in the Ceph Storage cluster some time after the gateway deletes the objects from the bucket index. The process of purging the deleted object data from the Ceph Storage cluster is known as Garbage Collection or GC.

To view the queue of objects awaiting garbage collection, execute the following:

1. `$ radosgw-admin gc list`
- 2.
3. **Note:** specify `--include-all` to list all entries, including unexpired

Garbage collection is a background activity that may execute continuously or during times of low loads, depending upon how the administrator configures the Ceph Object Gateway. By default, the Ceph Object Gateway conducts GC operations continuously.

Since GC operations are a normal part of Ceph Object Gateway operations, especially with object delete operations, objects eligible for garbage collection exist most of the time.

Some workloads may temporarily or permanently outpace the rate of garbage collection activity. This is especially true of delete-heavy workloads, where many objects get stored for a short period of time and then deleted. For these types of workloads, administrators can increase the priority of garbage collection operations relative to other operations with the following configuration parameters.

```
rgw gc max objs
```

#### Description

The maximum number of objects that may be handled by garbage collection in one garbage collection processing cycle. Please do not change this value after the first deployment.

#### Type

Integer

#### Default

```
32
```

```
rgw gc obj min wait
```

#### Description

The minimum wait time before a deleted object may be removed and handled by garbage collection processing.

#### Type

Integer

#### Default

```
2 * 3600
```

```
rgw gc processor max time
```

#### Description

The maximum time between the beginning of two consecutive garbage collection processing cycles.

#### Type

Integer

#### Default

`3600``rgw gc processor period`

#### Description

The cycle time for garbage collection processing.

#### Type

Integer

#### Default

`3600``rgw gc max concurrent io`

#### Description

The maximum number of concurrent IO operations that the RGW garbage collection thread will use when purging old data.

#### Type

Integer

#### Default

`10`

### Tuning Garbage Collection for Delete Heavy Workloads

As an initial step towards tuning Ceph Garbage Collection to be more aggressive the following options are suggested to be increased from their default configuration values:

`rgw gc max concurrent io = 20``rgw gc max trim chunk = 64`

#### NOTE

Modifying these values requires a restart of the RGW service.

Once these values have been increased from default please monitor for performance of the cluster during Garbage Collection to verify no adverse performance issues due to the increased values.

## Multisite Settings

New in version Jewel.

You may include the following settings in your Ceph configuration file under each `[client.radosgw.{instance-name}]` instance.

**rgw zone****Description**

The name of the zone for the gateway instance. If no zone is set, a cluster-wide default can be configured with the command `radosgw-admin zone default`.

**Type****String****Default****None****rgw zonegroup****Description**

The name of the zonegroup for the gateway instance. If no zonegroup is set, a cluster-wide default can be configured with the command `radosgw-admin zonegroup default`.

**Type****String****Default****None****rgw realm****Description**

The name of the realm for the gateway instance. If no realm is set, a cluster-wide default can be configured with the command `radosgw-admin realm default`.

**Type****String****Default****None****rgw run sync thread****Description**

If there are other zones in the realm to sync from, spawn threads to handle the sync of data and metadata.

**Type****Boolean**

**Default**`true``rgw data log window`**Description**

The data log entries window in seconds.

**Type****Integer****Default**`30``rgw data log changes size`**Description**

The number of in-memory entries to hold for the data changes log.

**Type****Integer****Default**`1000``rgw data log obj prefix`**Description**

The object name prefix for the data log.

**Type****String****Default**`data_log``rgw data log num shards`**Description**

The number of shards (objects) on which to keep the data changes log.

**Type****Integer****Default**

`128``rgw md log max shards`

#### Description

The maximum number of shards for the metadata log.

#### Type

Integer

#### Default

`64`

#### Important

The values of `rgw data log num shards` and `rgw md log max shards` should not be changed after sync has started.

## S3 Settings

---

`rgw s3 auth use ldap`

#### Description

Should S3 authentication use LDAP.

#### Type

Boolean

#### Default

`false`

## Swift Settings

---

`rgw enforce swift acls`

#### Description

Enforces the Swift Access Control List (ACL) settings.

#### Type

Boolean

#### Default

`true``rgw swift token expiration`

## Description

The time in seconds for expiring a Swift token.

## Type

Integer

## Default

```
24 * 3600
```

```
rgw swift url
```

## Description

The URL for the Ceph Object Gateway Swift API.

## Type

String

## Default

None

```
rgw swift url prefix
```

## Description

The URL prefix for the Swift API, to distinguish it from the S3 API endpoint. The default is `swift`, which makes the Swift API available at the URL

`http://host:port/swift/v1` (or `http://host:port/swift/v1/AUTH_%(tenant_id)s` if `rgw swift account in url` is enabled).

For compatibility, setting this configuration variable to the empty string causes the default `swift` to be used; if you do want an empty prefix, set this option to `/`.

## Warning

If you set this option to `/`, you must disable the S3 API by modifying `rgw enable apis` to exclude `s3`. It is not possible to operate radosgw with `rgw swift url prefix = /` and simultaneously support both the S3 and Swift APIs. If you do need to support both APIs without prefixes, deploy multiple radosgw instances to listen on different hosts (or ports) instead, enabling some for S3 and some for Swift.

## Default

```
swift
```

## Example

```
"/swift-testing"
```

```
rgw swift auth url
```

#### Description

Default URL for verifying v1 auth tokens (if not using internal Swift auth).

#### Type

String

#### Default

None

```
rgw swift auth entry
```

#### Description

The entry point for a Swift auth URL.

#### Type

String

#### Default

```
auth
```

```
rgw swift account in url
```

#### Description

Whether or not the Swift account name should be included in the Swift API URL.

If set to `false` (the default), then the Swift API will listen on a URL formed like `http://host:port/<rgw_swift_url_prefix>/v1`, and the account name (commonly a Keystone project UUID if radosgw is configured with [Keystone integration](#)) will be inferred from request headers.

If set to `true`, the Swift API URL will be

```
http://host:port/<rgw_swift_url_prefix>/v1/AUTH_<account_name> (or
http://host:port/<rgw_swift_url_prefix>/v1/AUTH_<keystone_project_id>) instead, and the Keystone
object-store endpoint must accordingly be configured to include the AUTH_%(tenant_id)s
suffix.
```

You **must** set this option to `true` (and update the Keystone service catalog) if you want radosgw to support publicly-readable containers and [temporary URLs](#).

#### Type

Boolean

#### Default

```
false
rgw swift versioning enabled
```

#### Description

Enables the Object Versioning of OpenStack Object Storage API. This allows clients to put the `X-Versions-Location` attribute on containers that should be versioned. The attribute specifies the name of container storing archived versions. It must be owned by the same user that the versioned container due to access control verification - ACLs are NOT taken into consideration. Those containers cannot be versioned by the S3 object versioning mechanism.

A slightly different attribute, `X-History-Location`, which is also understood by OpenStack Swift for handling `DELETE` operations, is currently not supported.

#### Type

Boolean

#### Default

```
false
rgw trust forwarded https
```

#### Description

When a proxy in front of radosgw is used for ssl termination, radosgw does not know whether incoming http connections are secure. Enable this option to trust the `Forwarded` and `X-Forwarded-Proto` headers sent by the proxy when determining whether the connection is secure. This is required for some features, such as server side encryption. (Never enable this setting if you do not have a trusted proxy in front of radosgw, or else malicious users will be able to set these headers in any request.)

#### Type

Boolean

#### Default

```
false
```

## Logging Settings

```
rgw log nonexistent bucket
```

#### Description

Enables Ceph Object Gateway to log a request for a non-existent bucket.

#### Type

Boolean

Default

false

rgw log object name

Description

The logging format for an object name. See manpage *date* for details about format specifiers.

Type

Date

Default

%Y-%m-%d-%H-%i-%n

rgw log object name utc

Description

Whether a logged object name includes a UTC time. If false , it uses the local time.

Type

Boolean

Default

false

rgw usage max shards

Description

The maximum number of shards for usage logging.

Type

Integer

Default

32

rgw usage max user shards

Description

The maximum number of shards used for a single user's usage logging.

Type

## Integer

### Default

1

```
rgw enable ops log
```

### Description

Enable logging for each successful Ceph Object Gateway operation.

### Type

Boolean

### Default

false

```
rgw enable usage log
```

### Description

Enable the usage log.

### Type

Boolean

### Default

false

```
rgw ops log rados
```

### Description

Whether the operations log should be written to the Ceph Storage Cluster backend.

### Type

Boolean

### Default

true

```
rgw ops log socket path
```

### Description

The Unix domain socket for writing operations logs.

### Type

String

**Default**

None

```
rgw ops log data backlog
```

**Description**

The maximum data backlog data size for operations logs written to a Unix domain socket.

Type

Integer

**Default**

```
5 << 20
```

```
rgw usage log flush threshold
```

**Description**

The number of dirty merged entries in the usage log before flushing synchronously.

Type

Integer

**Default**

1024

```
rgw usage log tick interval
```

**Description**

Flush pending usage log data every **n** seconds.

Type

Integer

**Default**

```
30
```

```
rgw log http headers
```

**Description**

Comma-delimited list of HTTP headers to include with ops log entries. Header names are case insensitive, and use the full header name with words separated by underscores.

Type

Type

Default

None

Example

```
"http_x_forwarded_for, http_x_special_k"
```

```
rgw intent log object name
```

Description

The logging format for the intent log object name. See manpage *date* for details about format specifiers.

Type

Date

Default

```
%Y-%m-%d-%i-%n
```

```
rgw intent log object name utc
```

Description

Whether the intent log object name includes a UTC time. If `false`, it uses the local time.

Type

Boolean

Default

```
false
```

## Keystone Settings

```
rgw keystone url
```

Description

The URL for the Keystone server.

Type

String

Default

None

```
rgw keystone api version
```

Description

The version (2 or 3) of OpenStack Identity API that should be used for communication with the Keystone server.

Type

Integer

Default

```
2
```

```
rgw keystone admin domain
```

Description

The name of OpenStack domain with admin privilege when using OpenStack Identity API v3.

Type

String

Default

None

```
rgw keystone admin project
```

Description

The name of OpenStack project with admin privilege when using OpenStack Identity API v3. If left unspecified, value of `rgw keystone admin tenant` will be used instead.

Type

String

Default

None

```
rgw keystone admin token
```

Description

The Keystone admin token (shared secret). In Ceph RadosGW authentication with the admin token has priority over authentication with the admin credentials (`rgw keystone admin user` , `rgw keystone admin password` , `rgw keystone admin tenant` , `rgw keystone admin project` , `rgw keystone admin domain` ). The Keystone admin token has been deprecated, but can be used to

integrate with older environments. Prefer `rgw keystone admin token path` to avoid exposing the token.

Type

String

Default

None

`rgw keystone admin token path`

Description

Path to a file containing the Keystone admin token (shared secret). In Ceph RadosGW authentication with the admin token has priority over authentication with the admin credentials (`rgw keystone admin user`, `rgw keystone admin password`, `rgw keystone admin tenant`, `rgw keystone admin project`, `rgw keystone admin domain`). The Keystone admin token has been deprecated, but can be used to integrate with older environments.

Type

String

Default

None

`rgw keystone admin tenant`

Description

The name of OpenStack tenant with admin privilege (Service Tenant) when using OpenStack Identity API v2

Type

String

Default

None

`rgw keystone admin user`

Description

The name of OpenStack user with admin privilege for Keystone authentication (Service User) when OpenStack Identity API v2

Type

String

Default

None

```
rgw keystone admin password
```

Description

The password for OpenStack admin user when using OpenStack Identity API v2. Prefer

```
rgw keystone admin password path
```

to avoid exposing the token.

Type

String

Default

None

```
rgw keystone admin password path
```

Description

Path to a file containing the password for OpenStack admin user when using OpenStack Identity API v2.

Type

String

Default

None

```
rgw keystone accepted roles
```

Description

The roles requires to serve requests.

Type

String

Default

```
Member, admin
```

```
rgw keystone token cache size
```

Description

The maximum number of entries in each Keystone token cache.

Type

Integer

Default

10000

rgw keystone revocation interval

Description

The number of seconds between token revocation checks.

Type

Integer

Default

15 \* 60

rgw keystone verify ssl

Description

Verify SSL certificates while making token requests to keystone.

Type

Boolean

Default

true

## Server-side encryption Settings

rgw crypt s3 kms backend

Description

Where the SSE-KMS encryption keys are stored. Supported KMS systems are OpenStack Barbican ( `barbican` , the default) and HashiCorp Vault ( `vault` ).

Type

String

Default

None

## Barbican Settings

```
rgw barbican url
```

#### Description

The URL for the Barbican server.

#### Type

String

#### Default

None

```
rgw keystone barbican user
```

#### Description

The name of the OpenStack user with access to the [Barbican](#) secrets used for [Encryption](#).

#### Type

String

#### Default

None

```
rgw keystone barbican password
```

#### Description

The password associated with the [Barbican](#) user.

#### Type

String

#### Default

None

```
rgw keystone barbican tenant
```

#### Description

The name of the OpenStack tenant associated with the [Barbican](#) user when using OpenStack Identity API v2.

#### Type

String

#### Default

None

```
rgw keystone barbican project
```

Description

The name of the OpenStack project associated with the `Barbican` user when using OpenStack Identity API v3.

Type

String

Default

None

```
rgw keystone barbican domain
```

Description

The name of the OpenStack domain associated with the `Barbican` user when using OpenStack Identity API v3.

Type

String

Default

None

## HashiCorp Vault Settings

```
rgw crypt vault auth
```

Description

Type of authentication method to be used. The only method currently supported is `token`.

Type

String

Default

```
token
```

```
rgw crypt vault token file
```

Description

If authentication method is `token`, provide a path to the token file, which should be

readable only by Rados Gateway.

Type

String

Default

None

```
rgw crypt vault addr
```

Description

Vault server base address, e.g. `http://vaultserver:8200`.

Type

String

Default

None

```
rgw crypt vault prefix
```

Description

The Vault secret URL prefix, which can be used to restrict access to a particular subset of the secret space, e.g. `/v1/secret/data`.

Type

String

Default

None

```
rgw crypt vault secret engine
```

Description

Vault Secret Engine to be used to retrieve encryption keys: choose between kv-v2, transit.

Type

String

Default

None

```
rgw crypt vault namespace
```

## Description

If set, Vault Namespace provides tenant isolation for teams and individuals on the same Vault Enterprise instance, e.g. `acme/tenant1`

Type

String

Default

None

## QoS settings

New in version Nautilus.

The `civetweb` frontend has a threading model that uses a thread per connection and hence automatically throttled by `rgw thread pool size` configurable when it comes to accepting connections. The `beast` frontend is not restricted by the thread pool size when it comes to accepting new connections, so a scheduler abstraction is introduced in Nautilus release which for supporting ways for scheduling requests in the future.

Currently the scheduler defaults to a throttler which throttles the active connections to a configured limit. QoS based on `mClock` is currently in an *experimental* phase and not recommended for production yet. Current implementation of `dmclock_client` op queue divides RGW Ops on admin, auth (swift auth, sts) metadata & data requests.

`rgw max concurrent requests`

## Description

Maximum number of concurrent HTTP requests that the beast frontend will process. Tuning this can help to limit memory usage under heavy load.

Type

Integer

Default

1024

`rgw scheduler type`

## Description

The type of RGW Scheduler to use. Valid values are `throttler`, `dmclock`. Currently defaults to `throttler` which throttles beast frontend requests. `dmclock` is *experimental* and will need the `experimental` flag set

The options below are to tune the experimental `dmclock` scheduler. For some further

reading on dmclock, see [QoS Based on mClock](#). op\_class for the flags below is one of admin, auth, metadata or data.

`rgw_dmclock_<op_class>_res`

Description

The mclock reservation for op\_class requests

Type

float

Default

100.0

`rgw_dmclock_<op_class>_wgt`

Description

The mclock weight for op\_class requests

Type

float

Default

1.0

`rgw_dmclock_<op_class>_lim`

Description

The mclock limit for op\_class requests

Type

float

Default

0.0

# Admin Guide

Once you have your Ceph Object Storage service up and running, you may administer the service with user management, access controls, quotas and usage tracking among other features.

## User Management

Ceph Object Storage user management refers to users of the Ceph Object Storage service (i.e., not the Ceph Object Gateway as a user of the Ceph Storage Cluster). You must create a user, access key and secret to enable end users to interact with Ceph Object Gateway services.

There are two user types:

- **User:** The term ‘user’ reflects a user of the S3 interface.
- **Subuser:** The term ‘subuser’ reflects a user of the Swift interface. A subuser is associated to a user .

You can create, modify, view, suspend and remove users and subusers. In addition to user and subuser IDs, you may add a display name and an email address for a user. You can specify a key and secret, or generate a key and secret automatically. When generating or specifying keys, note that user IDs correspond to an S3 key type and subuser IDs correspond to a swift key type. Swift keys also have access levels of `read` , `write` , `readwrite` and `full` .

## Create a User

To create a user (S3 interface), execute the following:

```
1. radosgw-admin user create --uid={username} --display-name="{display-name}" [--email={email}]
```

For example:

```
1. radosgw-admin user create --uid=johndoe --display-name="John Doe" --email=john@example.com
```

```
1. { "user_id": "johndoe",
2.   "display_name": "John Doe",
3.   "email": "john@example.com",
4.   "suspended": 0,
5.   "max_buckets": 1000,
6.   "subusers": [],
7.   "keys": [
8.     { "user": "johndoe",
```

```

9.      "access_key": "11BS02LGFB6AL6H1ADMW",
10.     "secret_key": "vzCEkuryfn060dfee4fgQPqFrncKEIkh3Zcd0ANY"}],
11.     "swift_keys": [],
12.     "caps": [],
13.     "op_mask": "read, write, delete",
14.     "default_placement": "",
15.     "placement_tags": [],
16.     "bucket_quota": { "enabled": false,
17.       "max_size_kb": -1,
18.       "max_objects": -1},
19.     "user_quota": { "enabled": false,
20.       "max_size_kb": -1,
21.       "max_objects": -1},
22.     "temp_url_keys": []}

```

Creating a user also creates an `access_key` and `secret_key` entry for use with any S3 API-compatible client.

### Important

Check the key output. Sometimes `radosgw-admin` generates a JSON escape (`\`) character, and some clients do not know how to handle JSON escape characters. Remedies include removing the JSON escape character (`\`), encapsulating the string in quotes, regenerating the key and ensuring that it does not have a JSON escape character or specify the key and secret manually.

## Create a Subuser

To create a subuser (Swift interface) for the user, you must specify the user ID (`--uid={username}`), a subuser ID and the access level for the subuser.

```
1. radosgw-admin subuser create --uid={uid} --subuser={uid} --access=[ read | write | readwrite | full ]
```

For example:

```
1. radosgw-admin subuser create --uid=johndoe --subuser=johndoe:swift --access=full
```

### Note

`full` is not `readwrite`, as it also includes the access control policy.

```

1. { "user_id": "johndoe",
2.   "display_name": "John Doe",
3.   "email": "john@example.com",
4.   "suspended": 0,
5.   "max_buckets": 1000,
6.   "subusers": [
7.     { "id": "johndoe:swift",
8.       "permissions": "full-control"}],

```

```

9.   "keys": [
10.     { "user": "johndoe",
11.       "access_key": "11BS02LGFB6AL6H1ADMW",
12.       "secret_key": "vzCEkuryfn060dfee4fgQPqFrncKEIkh3Zcd0ANY"]},
13.   "swift_keys": [],
14.   "caps": [],
15.   "op_mask": "read, write, delete",
16.   "default_placement": "",
17.   "placement_tags": [],
18.   "bucket_quota": { "enabled": false,
19.     "max_size_kb": -1,
20.     "max_objects": -1},
21.   "user_quota": { "enabled": false,
22.     "max_size_kb": -1,
23.     "max_objects": -1},
24.   "temp_url_keys": []}

```

## Get User Info

To get information about a user, you must specify `user info` and the user ID (`--uid={username}`) .

```
1. radosgw-admin user info --uid=johndoe
```

## Modify User Info

To modify information about a user, you must specify the user ID (`--uid={username}`) and the attributes you want to modify. Typical modifications are to keys and secrets, email addresses, display names and access levels. For example:

```
1. radosgw-admin user modify --uid=johndoe --display-name="John E. Doe"
```

To modify subuser values, specify `subuser modify`, user ID and the subuser ID. For example:

```
1. radosgw-admin subuser modify --uid=johndoe --subuser=johndoe.swift --access=full
```

## User Enable/Suspend

When you create a user, the user is enabled by default. However, you may suspend user privileges and re-enable them at a later time. To suspend a user, specify `user suspend` and the user ID.

```
1. radosgw-admin user suspend --uid=johndoe
```

To re-enable a suspended user, specify `user enable` and the user ID.

```
1. radosgw-admin user enable --uid=johndoe
```

## Note

Disabling the user disables the subuser.

## Remove a User

When you remove a user, the user and subuser are removed from the system. However, you may remove just the subuser if you wish. To remove a user (and subuser), specify `user rm` and the user ID.

```
1. radosgw-admin user rm --uid=johndoe
```

To remove the subuser only, specify `subuser rm` and the subuser ID.

```
1. radosgw-admin subuser rm --subuser=johndoe:swift
```

Options include:

- **Purge Data:** The `--purge-data` option purges all data associated to the UID.
- **Purge Keys:** The `--purge-keys` option purges all keys associated to the UID.

## Remove a Subuser

When you remove a sub user, you are removing access to the Swift interface. The user will remain in the system. To remove the subuser, specify `subuser rm` and the subuser ID.

```
1. radosgw-admin subuser rm --subuser=johndoe:swift
```

Options include:

- **Purge Keys:** The `--purge-keys` option purges all keys associated to the UID.

## Add / Remove a Key

Both users and subusers require the key to access the S3 or Swift interface. To use S3, the user needs a key pair which is composed of an access key and a secret key. On the other hand, to use Swift, the user typically needs a secret key (password), and use it together with the associated user ID. You may create a key and either specify or generate the access key and/or secret key. You may also remove a key. Options include:

- `--key-type=<type>` specifies the key type. The options are: s3, swift

- `--access-key=<key>` manually specifies an S3 access key.
- `--secret-key=<key>` manually specifies a S3 secret key or a Swift secret key.
- `--gen-access-key` automatically generates a random S3 access key.
- `--gen-secret` automatically generates a random S3 secret key or a random Swift secret key.

An example how to add a specified S3 key pair for a user.

```
1. radosgw-admin key create --uid=foo --key-type=s3 --access-key fooAccessKey --secret-key fooSecretKey
```

```
1. { "user_id": "foo",
2.   "rados_uid": 0,
3.   "display_name": "foo",
4.   "email": "foo@example.com",
5.   "suspended": 0,
6.   "keys": [
7.     { "user": "foo",
8.       "access_key": "fooAccessKey",
9.       "secret_key": "fooSecretKey"}],
10. }
```

Note that you may create multiple S3 key pairs for a user.

To attach a specified swift secret key for a subuser.

```
1. radosgw-admin key create --subuser=foo:bar --key-type=swift --secret-key barSecret
```

```
1. { "user_id": "foo",
2.   "rados_uid": 0,
3.   "display_name": "foo",
4.   "email": "foo@example.com",
5.   "suspended": 0,
6.   "subusers": [
7.     { "id": "foo:bar",
8.       "permissions": "full-control"}],
9.   "swift_keys": [
10.     { "user": "foo:bar",
11.       "secret_key": "asfghjghghmgm"}]
```

Note that a subuser can have only one swift secret key.

Subusers can also be used with S3 APIs if the subuser is associated with a S3 key pair.

```
1. radosgw-admin key create --subuser=foo:bar --key-type=s3 --access-key barAccessKey --secret-key barSecretKey
```

```

1. { "user_id": "foo",
2.   "rados_uid": 0,
3.   "display_name": "foo",
4.   "email": "foo@example.com",
5.   "suspended": 0,
6.   "subusers": [
7.     { "id": "foo:bar",
8.       "permissions": "full-control"}],
9.   "keys": [
10.    { "user": "foo:bar",
11.      "access_key": "barAccessKey",
12.      "secret_key": "barSecretKey"}],
13. }

```

To remove a S3 key pair, specify the access key.

```
1. radosgw-admin key rm --uid=foo --key-type=s3 --access-key=fooAccessKey
```

To remove the swift secret key.

```
1. radosgw-admin key rm --subuser=foo:bar --key-type=swift
```

## Add / Remove Admin Capabilities

The Ceph Storage Cluster provides an administrative API that enables users to execute administrative functions via the REST API. By default, users do NOT have access to this API. To enable a user to exercise administrative functionality, provide the user with administrative capabilities.

To add administrative capabilities to a user, execute the following:

```
1. radosgw-admin caps add --uid={uid} --caps={caps}
```

You can add read, write or all capabilities to users, buckets, metadata and usage (utilization). For example:

```
1. --caps="[users|buckets|metadata|usage|zone]=[*|read|write|read, write]"
```

For example:

```
1. radosgw-admin caps add --uid=johndoe --caps="users=*;buckets=*"
```

To remove administrative capabilities from a user, execute the following:

```
1. radosgw-admin caps rm --uid=johndoe --caps={caps}
```

# Quota Management

The Ceph Object Gateway enables you to set quotas on users and buckets owned by users. Quotas include the maximum number of objects in a bucket and the maximum storage size a bucket can hold.

- **Bucket:** The `--bucket` option allows you to specify a quota for buckets the user owns.
- **Maximum Objects:** The `--max-objects` setting allows you to specify the maximum number of objects. A negative value disables this setting.
- **Maximum Size:** The `--max-size` option allows you to specify a quota size in B/K/M/G/T, where B is the default. A negative value disables this setting.
- **Quota Scope:** The `--quota-scope` option sets the scope for the quota. The options are `bucket` and `user`. Bucket quotas apply to buckets a user owns. User quotas apply to a user.

## Set User Quota

Before you enable a quota, you must first set the quota parameters. For example:

```
1. radosgw-admin quota set --quota-scope=user --uid=<uid> [--max-objects=<num objects>] [--max-size=<max size>]
```

For example:

```
1. radosgw-admin quota set --quota-scope=user --uid=johndoe --max-objects=1024 --max-size=1024B
```

A negative value for num objects and / or max size means that the specific quota attribute check is disabled.

## Enable/Disable User Quota

Once you set a user quota, you may enable it. For example:

```
1. radosgw-admin quota enable --quota-scope=user --uid=<uid>
```

You may disable an enabled user quota. For example:

```
1. radosgw-admin quota disable --quota-scope=user --uid=<uid>
```

## Set Bucket Quota

Bucket quotas apply to the buckets owned by the specified `uid`. They are independent

of the user.

```
1. radosgw-admin quota set --uid=<uid> --quota-scope=bucket [--max-objects=<num objects>] [--max-size=<max size>]
```

A negative value for num objects and / or max size means that the specific quota attribute check is disabled.

## Enable/Disable Bucket Quota

Once you set a bucket quota, you may enable it. For example:

```
1. radosgw-admin quota enable --quota-scope=bucket --uid=<uid>
```

You may disable an enabled bucket quota. For example:

```
1. radosgw-admin quota disable --quota-scope=bucket --uid=<uid>
```

## Get Quota Settings

You may access each user's quota settings via the user information API. To read user quota setting information with the CLI interface, execute the following:

```
1. radosgw-admin user info --uid=<uid>
```

## Update Quota Stats

Quota stats get updated asynchronously. You can update quota statistics for all users and all buckets manually to retrieve the latest quota stats.

```
1. radosgw-admin user stats --uid=<uid> --sync-stats
```

## Get User Usage Stats

To see how much of the quota a user has consumed, execute the following:

```
1. radosgw-admin user stats --uid=<uid>
```

### Note

You should execute `radosgw-admin user stats` with the `--sync-stats` option to receive the latest data.

## Default Quotas

You can set default quotas in the config. These defaults are used when creating a new user and have no effect on existing users. If the relevant default quota is set in config, then that quota is set on the new user, and that quota is enabled. See `rgw bucket default quota max objects`, `rgw bucket default quota max size`, `rgw user default quota max objects`, and `rgw user default quota max size` in [Ceph Object Gateway Config Reference](#)

## Quota Cache

Quota statistics are cached on each RGW instance. If there are multiple instances, then the cache can keep quotas from being perfectly enforced, as each instance will have a different view of quotas. The options that control this are `rgw bucket quota ttl`, `rgw user quota bucket sync interval` and `rgw user quota sync interval`. The higher these values are, the more efficient quota operations are, but the more out-of-sync multiple instances will be. The lower these values are, the closer to perfect enforcement multiple instances will achieve. If all three are 0, then quota caching is effectively disabled, and multiple instances will have perfect quota enforcement. See [Ceph Object Gateway Config Reference](#)

## Reading / Writing Global Quotas

You can read and write global quota settings in the period configuration. To view the global quota settings:

```
1. radosgw-admin global quota get
```

The global quota settings can be manipulated with the `global quota` counterparts of the `quota set`, `quota enable`, and `quota disable` commands.

```
1. radosgw-admin global quota set --quota-scope bucket --max-objects 1024
2. radosgw-admin global quota enable --quota-scope bucket
```

### Note

In a multisite configuration, where there is a realm and period present, changes to the global quotas must be committed using `period update --commit`. If there is no period present, the rados gateway(s) must be restarted for the changes to take effect.

## Usage

The Ceph Object Gateway logs usage for each user. You can track user usage within date ranges too.

- Add `rgw enable usage log = true` in [client.rgw] section of ceph.conf and restart the radosgw service.

Options include:

- **Start Date:** The `--start-date` option allows you to filter usage stats from a particular start date (**format:** `yyyy-mm-dd[HH:MM:SS]` ).
- **End Date:** The `--end-date` option allows you to filter usage up to a particular date (**format:** `yyyy-mm-dd[HH:MM:SS]` ).
- **Log Entries:** The `--show-log-entries` option allows you to specify whether or not to include log entries with the usage stats (options: `true` | `false` ).

#### Note

You may specify time with minutes and seconds, but it is stored with 1 hour resolution.

## Show Usage

To show usage statistics, specify the `usage show`. To show usage for a particular user, you must specify a user ID. You may also specify a start date, end date, and whether or not to show log entries.:

```
1. radosgw-admin usage show --uid=johndoe --start-date=2012-03-01 --end-date=2012-04-01
```

You may also show a summary of usage information for all users by omitting a user ID.

```
1. radosgw-admin usage show --show-log-entries=false
```

## Trim Usage

With heavy use, usage logs can begin to take up storage space. You can trim usage logs for all users and for specific users. You may also specify date ranges for trim operations.

```
1. radosgw-admin usage trim --start-date=2010-01-01 --end-date=2010-12-31
2. radosgw-admin usage trim --uid=johndoe
3. radosgw-admin usage trim --uid=johndoe --end-date=2013-12-31
```

# Ceph Object Gateway S3 API

Ceph supports a RESTful API that is compatible with the basic data access model of the [Amazon S3 API](#).

## API

- [Common](#)
- [Authentication](#)
- [Service Ops](#)
- [Bucket Ops](#)
- [Object Ops](#)
- [C++](#)
- [C#](#)
- [Java](#)
- [Perl](#)
- [PHP](#)
- [Python](#)
- [Ruby AWS::SDK Examples \(aws-sdk gem ~>2\)](#)
- [Ruby AWS::S3 Examples \(aws-s3 gem\)](#)

## Features Support

The following table describes the support status for current Amazon S3 functional features:

Feature	Status	Remarks
<b>List Buckets</b>	Supported	
<b>Delete Bucket</b>	Supported	
<b>Create Bucket</b>	Supported	Different set of canned ACLs
<b>Bucket Lifecycle</b>	Supported	
<b>Policy (Buckets, Objects)</b>	Supported	ACLs & bucket policies are supported
<b>Bucket Website</b>	Supported	
<b>Bucket ACLs (Get, Put)</b>	Supported	Different set of canned ACLs
<b>Bucket Location</b>	Supported	

<b>Bucket Notification</b>	Supported	See <a href="#">S3 Notification Compatibility</a>
<b>Bucket Object Versions</b>	Supported	
<b>Get Bucket Info (HEAD)</b>	Supported	
<b>Bucket Request Payment</b>	Supported	
<b>Put Object</b>	Supported	
<b>Delete Object</b>	Supported	
<b>Get Object</b>	Supported	
<b>Object ACLs (Get, Put)</b>	Supported	
<b>Get Object Info (HEAD)</b>	Supported	
<b>POST Object</b>	Supported	
<b>Copy Object</b>	Supported	
<b>Multipart Uploads</b>	Supported	
<b>Object Tagging</b>	Supported	See <a href="#">Object Related Operations for Policy verbs</a>
<b>Bucket Tagging</b>	Supported	
<b>Storage Class</b>	Supported	See <a href="#">Storage Classes</a>

## Unsupported Header Fields

The following common request header fields are not supported:

Name	Type
<b>x-amz-security-token</b>	Request
<b>Server</b>	Response
<b>x-amz-delete-marker</b>	Response
<b>x-amz-id-2</b>	Response
<b>x-amz-version-id</b>	Response



# Common Entities

## Bucket and Host Name

There are two different modes of accessing the buckets. The first (preferred) method identifies the bucket as the top-level directory in the URI.

1. GET /mybucket HTTP/1.1
2. Host: cname.domain.com

The second method identifies the bucket via a virtual bucket host name. For example:

1. GET / HTTP/1.1
2. Host: mybucket.cname.domain.com

To configure virtual hosted buckets, you can either set `rgw_dns_name = cname.domain.com` in `ceph.conf`, or add `cname.domain.com` to the list of `hostnames` in your zonegroup configuration. See [Ceph Object Gateway - Multisite Configuration](#) for more on zonegroups.

### Tip

We prefer the first method, because the second method requires expensive domain certification and DNS wild cards.

## Common Request Headers

Request Header	Description
<code>CONTENT_LENGTH</code>	Length of the request body.
<code>DATE</code>	Request time and date (in UTC).
<code>HOST</code>	The name of the host server.
<code>AUTHORIZATION</code>	Authorization token.

## Common Response Status

HTTP Status	Response Code
<code>100</code>	Continue

200	Success
201	Created
202	Accepted
204	NoContent
206	Partial content
304	NotModified
400	InvalidArgumentException
400	InvalidDigest
400	BadDigest
400	InvalidBucketName
400	InvalidObjectName
400	UnresolvableGrantByEmailAddress
400	InvalidPart
400	InvalidPartOrder
400	RequestTimeout
400	EntityTooLarge
403	AccessDenied
403	UserSuspended
403	RequestTimeTooSkewed
404	NoSuchKey
404	NoSuchBucket
404	NoSuchUpload
405	MethodNotAllowed
408	RequestTimeout

409	BucketAlreadyExists
409	BucketNotEmpty
411	MissingContentLength
412	PreconditionFailed
416	InvalidRange
422	UnprocessableEntity
500	InternalError

# Authentication and ACLs

Requests to the RADOS Gateway (RGW) can be either authenticated or unauthenticated. RGW assumes unauthenticated requests are sent by an anonymous user. RGW supports canned ACLs.

## Authentication

Authenticating a request requires including an access key and a Hash-based Message Authentication Code (HMAC) in the request before it is sent to the RGW server. RGW uses an S3-compatible authentication approach.

1. `HTTP/1.1`
2. `PUT /buckets/bucket/object.mpeg`
3. `Host: cname.domain.com`
4. `Date: Mon, 2 Jan 2012 00:01:01 +0000`
5. `Content-Encoding: mpeg`
6. `Content-Length: 9999999`
- 7.
8. `Authorization: AWS {access-key}:{hash-of-header-and-secret}`

In the foregoing example, replace `{access-key}` with the value for your access key ID followed by a colon (`:`). Replace `{hash-of-header-and-secret}` with a hash of the header string and the secret corresponding to the access key ID.

To generate the hash of the header string and secret, you must:

1. Get the value of the header string.
2. Normalize the request header string into canonical form.
3. Generate an HMAC using a SHA-1 hashing algorithm. See [RFC 2104](#) and [HMAC](#) for details.
4. Encode the `hmac` result as base-64.

To normalize the header into canonical form:

1. Get all fields beginning with `x-amz-`.
2. Ensure that the fields are all lowercase.
3. Sort the fields lexicographically.
4. Combine multiple instances of the same field name into a single field and separate the field values with a comma.
5. Replace white space and line breaks in field values with a single space.

6. Remove white space before and after colons.
7. Append a new line after each field.
8. Merge the fields back into the header.

Replace the `{hash-of-header-and-secret}` with the base-64 encoded HMAC string.

## Authentication against OpenStack Keystone

In a radosgw instance that is configured with authentication against OpenStack Keystone, it is possible to use Keystone as an authoritative source for S3 API authentication. To do so, you must set:

- the `rgw keystone` configuration options explained in [Integrating with OpenStack Keystone](#),
- `rgw s3 auth use keystone = true`.

In addition, a user wishing to use the S3 API must obtain an AWS-style access key and secret key. They can do so with the `openstack ec2 credentials create` command:

```

1. $ openstack --os-interface public ec2 credentials create
+-----+
2. |-----+
| Field      | Value
3. |
+-----+
4. |-----+
| access      | c921676aaabbccdeadbeef7e8b0eeb2c
5. |
| links      | {u'self': u'https://auth.example.com:5000/v3/users/7ecbeaffabbddeadbeefa23267ccb24/credentials/OS-'
6. EC2/c921676aaabbccdeadbeef7e8b0eeb2c'} |
| project_id | 5ed51981aab4679851adeadbeef6ebf7
7. |
| secret      | *****
8. |
| trust_id    | None
9. |
| user_id     | 7ecbeaffabbddeadbeefa23267cc24
10. |
+-----+
11. -----+

```

The thus-generated access and secret key can then be used for S3 API access to radosgw.

### Note

Consider that most production radosgw deployments authenticating against OpenStack Keystone are also set up for [RGW Multi-tenancy](#), for which special considerations apply with respect to S3 signed URLs and public read ACLs.

# Access Control Lists (ACLs)

RGW supports S3-compatible ACL functionality. An ACL is a list of access grants that specify which operations a user can perform on a bucket or on an object. Each grant has a different meaning when applied to a bucket versus applied to an object:

Permission	Bucket	Object
READ	Grantee can list the objects in the bucket.	Grantee can read the object.
WRITE	Grantee can write or delete objects in the bucket.	N/A
READ_ACP	Grantee can read bucket ACL.	Grantee can read the object ACL.
WRITE_ACP	Grantee can write bucket ACL.	Grantee can write to the object ACL.
FULL_CONTROL	Grantee has full permissions for object in the bucket.	Grantee can read or write to the object ACL.

Internally, S3 operations are mapped to ACL permissions thus:

Operation	Permission
s3:GetObject	READ
s3:GetObjectTorrent	READ
s3:GetObjectVersion	READ
s3:GetObjectVersionTorrent	READ
s3:GetObjectTagging	READ
s3:GetObjectVersionTagging	READ
s3>ListAllMyBuckets	READ
s3>ListBucket	READ
s3>ListBucketMultipartUploads	READ
s3>ListBucketVersions	READ
s3>ListMultipartUploadParts	READ
s3AbortMultipartUpload	WRITE

s3:CreateBucket	WRITE
s3>DeleteBucket	WRITE
s3>DeleteObject	WRITE
s3:s3DeleteObjectVersion	WRITE
s3:PutObject	WRITE
s3:PutObjectTagging	WRITE
s3:PutObjectVersionTagging	WRITE
s3:DeleteObjectTagging	WRITE
s3:DeleteObjectVersionTagging	WRITE
s3:RestoreObject	WRITE
s3:GetAccelerateConfiguration	READ_ACP
s3:GetBucketAcl	READ_ACP
s3:GetBucketCORS	READ_ACP
s3:GetBucketLocation	READ_ACP
s3:GetBucketLogging	READ_ACP
s3:GetBucketNotification	READ_ACP
s3:GetBucketPolicy	READ_ACP
s3:GetBucketRequestPayment	READ_ACP
s3:GetBucketTagging	READ_ACP
s3:GetBucketVersioning	READ_ACP
s3:GetBucketWebsite	READ_ACP
s3:GetLifecycleConfiguration	READ_ACP
s3:GetObjectAcl	READ_ACP
s3:GetObjectVersionAcl	READ_ACP
s3:GetReplicationConfiguration	READ_ACP
s3:DeleteBucketPolicy	WRITE_ACP

s3:DeleteBucketWebsite	WRITE_ACP
s3:DeleteReplicationConfiguration	WRITE_ACP
s3:PutAccelerateConfiguration	WRITE_ACP
s3:PutBucketAcl	WRITE_ACP
s3:PutBucketCORS	WRITE_ACP
s3:PutBucketLogging	WRITE_ACP
s3:PutBucketNotification	WRITE_ACP
s3:PutBucketPolicy	WRITE_ACP
s3:PutBucketRequestPayment	WRITE_ACP
s3:PutBucketTagging	WRITE_ACP
s3:PutPutBucketVersioning	WRITE_ACP
s3:PutBucketWebsite	WRITE_ACP
s3:PutLifecycleConfiguration	WRITE_ACP
s3:PutObjectAcl	WRITE_ACP
s3:PutObjectVersionAcl	WRITE_ACP
s3:PutReplicationConfiguration	WRITE_ACP

Some mappings, (e.g. `s3:CreateBucket` to `WRITE`) are not applicable to S3 operation, but are required to allow Swift and S3 to access the same resources when things like Swift user ACLs are in play. This is one of the many reasons that you should use S3 bucket policies rather than S3 ACLs when possible.

# Service Operations

## List Buckets

`GET /` returns a list of buckets created by the user making the request. `GET /` only returns buckets created by an authenticated user. You cannot make an anonymous request.

### Syntax

1. `GET / HTTP/1.1`
2. `Host: cname.domain.com`
- 3.
4. `Authorization: AWS {access-key}:{hash-of-header-and-secret}`

## Response Entities

Name	Type	Description
<code>Buckets</code>	Container	Container for list of buckets.
<code>Bucket</code>	Container	Container for bucket information.
<code>Name</code>	String	Bucket name.
<code>CreationDate</code>	Date	UTC time when the bucket was created.
<code>ListAllMyBucketsResult</code>	Container	A container for the result.
<code>Owner</code>	Container	A container for the bucket owner's <code>ID</code> and <code>DisplayName</code> .
<code>ID</code>	String	The bucket owner's ID.
<code>DisplayName</code>	String	The bucket owner's display name.

## Get Usage Stats

Gets usage stats per user, similar to the admin command [Get User Usage Stats](#).

### Syntax

```
1. GET /?usage HTTP/1.1
2. Host: cname.domain.com
3.
4. Authorization: AWS {access-key}:{hash-of-header-and-secret}
```

## Response Entities

Name	Type	Description
Summary	Container	Summary of total stats by user.
TotalBytes	Integer	Bytes used by user
TotalBytesRounded	Integer	Bytes rounded to the nearest 4k boundary
TotalEntries	Integer	Total object entries

# Bucket Operations

## PUT Bucket

Creates a new bucket. To create a bucket, you must have a user ID and a valid AWS Access Key ID to authenticate requests. You may not create buckets as an anonymous user.

### Constraints

In general, bucket names should follow domain name constraints.

- Bucket names must be unique.
- Bucket names cannot be formatted as IP address.
- Bucket names can be between 3 and 63 characters long.
- Bucket names must not contain uppercase characters or underscores.
- Bucket names must start with a lowercase letter or number.
- Bucket names must be a series of one or more labels. Adjacent labels are separated by a single period (.). Bucket names can contain lowercase letters, numbers, and hyphens. Each label must start and end with a lowercase letter or a number.

#### Note

The above constraints are relaxed if the option ‘rgw\_relaxed\_s3\_bucket\_names’ is set to true except that the bucket names must still be unique, cannot be formatted as IP address and can contain letters, numbers, periods, dashes and underscores for up to 255 characters long.

### Syntax

1. `PUT /{bucket} HTTP/1.1`
2. `Host: cname.domain.com`
3. `x-amz-acl: public-read-write`
- 4.
5. `Authorization: AWS {access-key}:{hash-of-header-and-secret}`

### Parameters

Name	Description	Valid Values	Required
------	-------------	--------------	----------

<code>x-amz-acl</code>	Canned ACLs.	<code>private</code> , <code>public-read</code> , <code>public-read-write</code> , <code>authenticated-read</code>	No
<code>x-amz-bucket-object-lock-enabled</code>	Enable object lock on bucket.	<code>true</code> , <code>false</code>	No

## Request Entities

Name	Type	Description
<code>CreateBucketConfiguration</code>	Container	A container for the bucket configuration.
<code>LocationConstraint</code>	String	A zonegroup api name, with optional <a href="#">S3 Bucket Placement</a>

## HTTP Response

If the bucket name is unique, within constraints and unused, the operation will succeed. If a bucket with the same name already exists and the user is the bucket owner, the operation will succeed. If the bucket name is already in use, the operation will fail.

HTTP Status	Status Code	Description
<code>409</code>	<code>BucketAlreadyExists</code>	Bucket already exists under different user's ownership.

## DELETE Bucket

Deletes a bucket. You can reuse bucket names following a successful bucket removal.

### Syntax

1. `DELETE /{bucket} HTTP/1.1`
2. `Host: cname.domain.com`
- 3.
4. `Authorization: AWS {access-key}:{hash-of-header-and-secret}`

## HTTP Response

HTTP Status	Status Code	Description
<code>204</code>	No Content	Bucket removed.

# GET Bucket

Returns a list of bucket objects.

## Syntax

1. GET `/{bucket}?max-keys=25` HTTP/1.1
2. Host: cname.domain.com

## Parameters

Name	Type	Description
<code>prefix</code>	String	Only returns objects that contain the specified prefix.
<code>delimiter</code>	String	The delimiter between the prefix and the rest of the object name.
<code>marker</code>	String	A beginning index for the list of objects returned.
<code>max-keys</code>	Integer	The maximum number of keys to return. Default is 1000.
<code>allow-unordered</code>	Boolean	Non-standard extension. Allows results to be returned unordered. Cannot be used with delimiter.

## HTTP Response

HTTP Status	Status Code	Description
<code>200</code>	OK	Buckets retrieved

## Bucket Response Entities

`GET /{bucket}` returns a container for buckets with the following fields.

Name	Type	Description
<code>ListBucketResult</code>	Entity	The container for the list of objects.
<code>Name</code>	String	The name of the bucket whose contents will be returned.

<code>Prefix</code>	String	A prefix for the object keys.
<code>Marker</code>	String	A beginning index for the list of objects returned.
<code>MaxKeys</code>	Integer	The maximum number of keys returned.
<code>Delimiter</code>	String	If set, objects with the same prefix will appear in the <code>CommonPrefixes</code> list.
<code>IsTruncated</code>	Boolean	If <code>true</code> , only a subset of the bucket's contents were returned.
<code>CommonPrefixes</code>	Container	If multiple objects contain the same prefix, they will appear in this list.

## Object Response Entities

The `ListBucketResult` contains objects, where each object is within a `Contents` container.

Name	Type	Description
<code>Contents</code>	Object	A container for the object.
<code>Key</code>	String	The object's key.
<code>LastModified</code>	Date	The object's last-modified date/time.
<code>ETag</code>	String	An MD-5 hash of the object. (entity tag)
<code>Size</code>	Integer	The object's size.
<code>StorageClass</code>	String	Should always return <code>STANDARD</code> .
<code>Type</code>	String	<code>Appendable</code> or <code>Normal</code> .

## Get Bucket Location

Retrieves the bucket's region. The user needs to be the bucket owner to call this. A bucket can be constrained to a region by providing `LocationConstraint` during a PUT request.

## Syntax

Add the `location` subresource to bucket resource as shown below

1. GET /{bucket}?location HTTP/1.1
2. Host: cname.domain.com
- 3.
4. Authorization: AWS {access-key}:{hash-of-header-and-secret}

## Response Entities

Name	Type	Description
LocationConstraint	String	The region where bucket resides, empty string for default region

## Get Bucket ACL

Retrieves the bucket access control list. The user needs to be the bucket owner or to have been granted `READ_ACP` permission on the bucket.

### Syntax

Add the `acl` subresource to the bucket request as shown below.

1. GET /{bucket}?acl HTTP/1.1
2. Host: cname.domain.com
- 3.
4. Authorization: AWS {access-key}:{hash-of-header-and-secret}

## Response Entities

Name	Type	Description
AccessControlPolicy	Container	A container for the response.
AccessControlList	Container	A container for the ACL information.
Owner	Container	A container for the bucket owner's <code>ID</code> and <code>DisplayName</code> .
ID	String	The bucket owner's ID.
DisplayName	String	The bucket owner's display name.
Grant	Container	A container for <code>Grantee</code> and <code>Permission</code> .
Grantee	Container	A container for the <code>DisplayName</code> and <code>ID</code> of the user receiving a grant of permission.

<code>Permission</code>	String	The permission given to the <code>Grantee</code> bucket.
-------------------------	--------	----------------------------------------------------------

## PUT Bucket ACL

Sets an access control to an existing bucket. The user needs to be the bucket owner or to have been granted `WRITE_ACP` permission on the bucket.

### Syntax

Add the `acl` subresource to the bucket request as shown below.

```
1. PUT /{bucket}?acl HTTP/1.1
```

### Request Entities

Name	Type	Description
<code>AccessControlPolicy</code>	Container	A container for the request.
<code>AccessControlList</code>	Container	A container for the ACL information.
<code>Owner</code>	Container	A container for the bucket owner's <code>ID</code> and <code>DisplayName</code> .
<code>ID</code>	String	The bucket owner's ID.
<code>DisplayName</code>	String	The bucket owner's display name.
<code>Grant</code>	Container	A container for <code>Grantee</code> and <code>Permission</code> .
<code>Grantee</code>	Container	A container for the <code>DisplayName</code> and <code>ID</code> of the user receiving a grant of permission.
<code>Permission</code>	String	The permission given to the <code>Grantee</code> bucket.

# List Bucket Multipart Uploads

`GET /?uploads` returns a list of the current in-progress multipart uploads-i.e., the application initiates a multipart upload, but the service hasn't completed all the uploads yet.

## Syntax

```
1. GET /{bucket}?uploads HTTP/1.1
```

## Parameters

You may specify parameters for `GET /{bucket}?uploads`, but none of them are required.

Name	Type	Description
<code>prefix</code>	String	Returns in-progress uploads whose keys contains the specified prefix.
<code>delimiter</code>	String	The delimiter between the prefix and the rest of the object name.
<code>key-marker</code>	String	The beginning marker for the list of uploads.
<code>max-keys</code>	Integer	The maximum number of in-progress uploads. The default is 1000.
<code>max-uploads</code>	Integer	The maximum number of multipart uploads. The range from 1-1000. The default is 1000.
<code>upload-id-marker</code>	String	Ignored if <code>key-marker</code> is not specified. Specifies the ID of first upload to list in lexicographical order at or following the ID.

## Response Entities

Name	Type	Description
<code>ListMultipartUploadsResult</code>	Container	A container for the results.
<code>ListMultipartUploadsResult.Prefix</code>	String	The prefix specified by the <code>prefix</code> request parameter (if any).
<code>Bucket</code>	String	The bucket that will receive the bucket contents.

<code>KeyMarker</code>	<code>String</code>	The key marker specified by the <code>key-marker</code> request parameter (if any).
<code>UploadIdMarker</code>	<code>String</code>	The marker specified by the <code>upload-id-marker</code> request parameter (if any).
<code>NextKeyMarker</code>	<code>String</code>	The key marker to use in a subsequent request if <code>IsTruncated</code> is <code>true</code> .
<code>NextUploadIdMarker</code>	<code>String</code>	The upload ID marker to use in a subsequent request if <code>IsTruncated</code> is <code>true</code> .
<code>MaxUploads</code>	<code>Integer</code>	The max uploads specified by the <code>max-uploads</code> request parameter.
<code>Delimiter</code>	<code>String</code>	If set, objects with the same prefix will appear in the <code>CommonPrefixes</code> list.
<code>IsTruncated</code>	<code>Boolean</code>	If <code>true</code> , only a subset of the bucket's upload contents were returned.
<code>Upload</code>	<code>Container</code>	A container for <code>Key</code> , <code>UploadId</code> , <code>InitiatorOwner</code> , <code>StorageClass</code> , and <code>Initiated</code> elements.
<code>Key</code>	<code>String</code>	The key of the object once the multipart upload is complete.
<code>UploadId</code>	<code>String</code>	The <code>ID</code> that identifies the multipart upload.
<code>Initiator</code>	<code>Container</code>	Contains the <code>ID</code> and <code>DisplayName</code> of the user who initiated the upload.
<code>DisplayName</code>	<code>String</code>	The initiator's display name.
<code>ID</code>	<code>String</code>	The initiator's ID.
<code>Owner</code>	<code>Container</code>	A container for the <code>ID</code> and <code>DisplayName</code> of the user who owns the uploaded object.
<code>StorageClass</code>	<code>String</code>	The method used to store the resulting object. <code>STANDARD</code> or <code>REDUCED_REDUNDANCY</code>
<code>Initiated</code>	<code>Date</code>	The date and time the user initiated the upload.
<code>CommonPrefixes</code>	<code>Container</code>	If multiple objects contain the same prefix, they will appear in this list.

<code>CommonPrefixes.Prefix</code>	String	The substring of the key after the prefix as defined by the <code>prefix</code> request parameter.
------------------------------------	--------	----------------------------------------------------------------------------------------------------

## ENABLE/SUSPEND BUCKET VERSIONING

`PUT /?versioning` This subresource set the versioning state of an existing bucket. To set the versioning state, you must be the bucket owner.

You can set the versioning state with one of the following values:

- Enabled : Enables versioning for the objects in the bucket, All objects added to the bucket receive a unique version ID.
- Suspended : Disables versioning for the objects in the bucket, All objects added to the bucket receive the version ID null.

If the versioning state has never been set on a bucket, it has no versioning state; a GET versioning request does not return a versioning state value.

## Syntax

```
1. PUT /{bucket}?versioning HTTP/1.1
```

## REQUEST ENTITIES

Name	Type	Description
<code>VersioningConfiguration</code>	Container	A container for the request.
<code>Status</code>	String	Sets the versioning state of the bucket. Valid Values: Suspended/Enabled

## PUT BUCKET OBJECT LOCK

Places an Object Lock configuration on the specified bucket. The rule specified in the Object Lock configuration will be applied by default to every new object placed in the specified bucket.

## Syntax

```
1. PUT /{bucket}?object-lock HTTP/1.1
```

## Request Entities

Name	Type	Description	Required
ObjectLockConfiguration	Container	A container for the request.	Yes
ObjectLockEnabled	String	Indicates whether this bucket has an Object Lock configuration enabled.	Yes
Rule	Container	The Object Lock rule in place for the specified bucket.	No
DefaultRetention	Container	The default retention period applied to new objects placed in the specified bucket.	No
Mode	String	The default Object Lock retention mode. Valid Values: GOVERNANCE/COMPLIANCE	Yes
Days	Integer	The number of days specified for the default retention period.	No
Years	Integer	The number of years specified for the default retention period.	No

## HTTP Response

If the bucket object lock is not enabled when creating the bucket, the operation will fail.

HTTP Status	Status Code	Description
400	MalformedXML	The XML is not well-formed
409	InvalidBucketState	The bucket object lock is not enabled

## GET BUCKET OBJECT LOCK

Gets the Object Lock configuration for a bucket. The rule specified in the Object Lock configuration will be applied by default to every new object placed in the specified bucket.

## Syntax

```
1. GET /{bucket}?object-lock HTTP/1.1
```

## Response Entities

Name	Type	Description	Required
ObjectLockConfiguration	Container	A container for the request.	Yes
ObjectLockEnabled	String	Indicates whether this bucket has an Object Lock configuration enabled.	Yes
Rule	Container	The Object Lock rule in place for the specified bucket.	No
DefaultRetention	Container	The default retention period applied to new objects placed in the specified bucket.	No
Mode	String	The default Object Lock retention mode. Valid Values: GOVERNANCE/COMPLIANCE	Yes
Days	Integer	The number of days specified for the default retention period.	No
Years	Integer	The number of years specified for the default retention period.	No

## Create Notification

Create a publisher for a specific bucket into a topic.

### Syntax

```
1. PUT /<bucket name>?notification HTTP/1.1
```

## Request Entities

Parameters are XML encoded in the body of the request, in the following format:

```
1. <NotificationConfiguration xmlns="http://s3.amazonaws.com/doc/2006-03-01/">
2.   <TopicConfiguration>
3.     <Id></Id>
4.     <Topic></Topic>
5.   <Event></Event>
```

```

6.      <Filter>
7.          <S3Key>
8.              <FilterRule>
9.                  <Name></Name>
10.                 <Value></Value>
11.             </FilterRule>
12.         </S3Key>
13.     <S3Metadata>
14.         <FilterRule>
15.             <Name></Name>
16.             <Value></Value>
17.         </FilterRule>
18.     </S3Metadata>
19.     <S3Tags>
20.         <FilterRule>
21.             <Name></Name>
22.             <Value></Value>
23.         </FilterRule>
24.     </S3Tags>
25.   </Filter>
26. </TopicConfiguration>
27. </NotificationConfiguration>

```

Name	Type	Description	Required
<code>NotificationConfiguration</code>	Container	Holding list of <code>TopicConfiguration</code> entities	Yes
<code>TopicConfiguration</code>	Container	Holding <code>Id</code> , <code>Topic</code> and list of <code>Event</code> entities	Yes
<code>Id</code>	String	Name of the notification	Yes
<code>Topic</code>	String	Topic ARN. Topic must be created beforehand	Yes
<code>Event</code>	String	List of supported events see: <a href="#">S3 Notification Compatibility</a> . Multiple <code>Event</code> entities can be used. If omitted, all events are handled	No
<code>Filter</code>	Container	Holding <code>S3Key</code> , <code>S3Metadata</code> and <code>S3Tags</code> entities	No
<code>S3Key</code>	Container	Holding a list of <code>FilterRule</code> entities, for filtering based on object key. At most, 3 entities may be in the list, with <code>Name</code> be <code>prefix</code> , <code>suffix</code> or <code>regex</code> . All filter rules in the list must match for the filter to match.	No

S3Metadata	Container	Holding a list of <code>FilterRule</code> entities, for filtering based on object metadata. All filter rules in the list must match the metadata defined on the object. However, the object still match if it has other metadata entries not listed in the filter.	No
S3Tags	Container	Holding a list of <code>FilterRule</code> entities, for filtering based on object tags. All filter rules in the list must match the tags defined on the object. However, the object still match if it has other tags not listed in the filter.	No
S3Key.FilterRule	Container	Holding <code>Name</code> and <code>Value</code> entities. <code>Name</code> would be: <code>prefix</code> , <code>suffix</code> or <code>regex</code> . The <code>Value</code> would hold the key prefix, key suffix or a regular expression for matching the key, accordingly.	Yes
S3Metadata.FilterRule	Container	Holding <code>Name</code> and <code>Value</code> entities. <code>Name</code> would be the name of the metadata attribute (e.g. <code>x-amz-meta-xxx</code> ). The <code>Value</code> would be the expected value for this attribute.	Yes
S3Tags.FilterRule	Container	Holding <code>Name</code> and <code>Value</code> entities. <code>Name</code> would be the tag key, and <code>Value</code> would be the tag value.	Yes

## HTTP Response

HTTP Status	Status Code	Description
400	MalformedXML	The XML is not well-formed
400	InvalidArgument	Missing Id; Missing/Invalid Topic ARN; Invalid Event
404	NoSuchBucket	The bucket does not exist
404	NoSuchKey	The topic does not exist

## Delete Notification

Delete a specific, or all, notifications from a bucket.

#### Note

- Notification deletion is an extension to the S3 notification API
- When the bucket is deleted, any notification defined on it is also deleted
- Deleting an unknown notification (e.g. double delete) is not considered an error

## Syntax

```
1. DELETE /bucket?notification[=<notification-id>] HTTP/1.1
```

## Parameters

Name	Type	Description
notification-id	String	Name of the notification. If not provided, all notifications on the bucket are deleted

## HTTP Response

HTTP Status	Status Code	Description
404	NoSuchBucket	The bucket does not exist

## Get/List Notification

Get a specific notification, or list all notifications configured on a bucket.

## Syntax

```
1. GET /bucket?notification[=<notification-id>] HTTP/1.1
```

## Parameters

Name	Type	Description
notification-id	String	Name of the notification. If not provided, all notifications on the bucket are listed

# Response Entities

Response is XML encoded in the body of the request, in the following format:

```

1. <NotificationConfiguration xmlns="http://s3.amazonaws.com/doc/2006-03-01/">
2.   <TopicConfiguration>
3.     <Id></Id>
4.     <Topic></Topic>
5.     <Event></Event>
6.     <Filter>
7.       <S3Key>
8.         <FilterRule>
9.           <Name></Name>
10.          <Value></Value>
11.        </FilterRule>
12.      </S3Key>
13.      <S3Metadata>
14.        <FilterRule>
15.          <Name></Name>
16.          <Value></Value>
17.        </FilterRule>
18.      </S3Metadata>
19.      <S3Tags>
20.        <FilterRule>
21.          <Name></Name>
22.          <Value></Value>
23.        </FilterRule>
24.      </S3Tags>
25.    </Filter>
26.  </TopicConfiguration>
27. </NotificationConfiguration>
```

Name	Type	Description	Required
NotificationConfiguration	Container	Holding list of TopicConfiguration entities	Yes
TopicConfiguration	Container	Holding Id , Topic and list of Event entities	Yes
Id	String	Name of the notification	Yes
Topic	String	Topic ARN	Yes
Event	String	Handled event. Multiple Event entities may exist	Yes
Filter	Container	Holding the filters configured for this notification	No

## HTTP Response

HTTP Status	Status Code	Description
404	NoSuchBucket	The bucket does not exist
404	NoSuchKey	The notification does not exist (if provided)

# Object Operations

## Put Object

Adds an object to a bucket. You must have write permissions on the bucket to perform this operation.

### Syntax

```
1. PUT /{bucket}/{object} HTTP/1.1
```

## Request Headers

Name	Description	Valid Values	Required
<b>content-md5</b>	A base64 encoded MD-5 hash of the message.	A string. No defaults or constraints.	No
<b>content-type</b>	A standard MIME type.	Any MIME type. Default: <code>binary/octet-stream</code>	No
<b>x-amz-meta-&lt;...&gt;</b>	User metadata. Stored with the object.	A string up to 8kb. No defaults.	No
<b>x-amz-acl</b>	A canned ACL.	<code>private</code> , <code>public-read</code> , <code>public-read-write</code> , <code>authenticated-read</code>	No

## Copy Object

To copy an object, use `PUT` and specify a destination bucket and the object name.

### Syntax

```
1. PUT /{dest-bucket}/{dest-object} HTTP/1.1
2. x-amz-copy-source: {source-bucket}/{source-object}
```

## Request Headers

Name	Description	Valid Values	Required
<b>x-amz-copy-source</b>	The source bucket name + object name.	{bucket}/{obj}	Yes

<b>x-amz-acl</b>	A canned ACL.	<code>private</code> , <code>public-read</code> , <code>public-read-write</code> , <code>authenticated-read</code>	No
<b>x-amz-copy-if-modified-since</b>	Copies only if modified since the timestamp.	Timestamp	No
<b>x-amz-copy-if-unmodified-since</b>	Copies only if unmodified since the timestamp.	Timestamp	No
<b>x-amz-copy-if-match</b>	Copies only if object ETag matches ETag.	Entity Tag	No
<b>x-amz-copy-if-none-match</b>	Copies only if object ETag doesn't match.	Entity Tag	No

## Response Entities

Name	Type	Description
<b>CopyObjectResult</b>	Container	A container for the response elements.
<b>LastModified</b>	Date	The last modified date of the source object.
<b>Etag</b>	String	The ETag of the new object.

## Remove Object

Removes an object. Requires WRITE permission set on the containing bucket.

### Syntax

```
1. DELETE /{bucket}/{object} HTTP/1.1
```

## Get Object

Retrieves an object from a bucket within RADOS.

### Syntax

```
1. GET /{bucket}/{object} HTTP/1.1
```

## Request Headers

Name	Description	Valid Values	Required
<b>range</b>	The range of the object to retrieve.	Range: bytes=beginbyte-endbyte	No
<b>if-modified-since</b>	Gets only if modified since the timestamp.	Timestamp	No
<b>if-unmodified-since</b>	Gets only if not modified since the timestamp.	Timestamp	No
<b>if-match</b>	Gets only if object ETag matches ETag.	Entity Tag	No
<b>if-none-match</b>	Gets only if object ETag matches ETag.	Entity Tag	No

## Response Headers

Name	Description
<b>Content-Range</b>	Data range, will only be returned if the range header field was specified in the request

## Get Object Info

Returns information about object. This request will return the same header information as with the Get Object request, but will include the metadata only, not the object data payload.

## Syntax

```
1. HEAD /{bucket}/{object} HTTP/1.1
```

## Request Headers

Name	Description	Valid Values	Required
<b>range</b>	The range of the object to retrieve.	Range: bytes=beginbyte-endbyte	No

<b>if-modified-since</b>	Gets only if modified since the timestamp.	Timestamp	No
<b>if-unmodified-since</b>	Gets only if not modified since the timestamp.	Timestamp	No
<b>if-match</b>	Gets only if object ETag matches ETag.	Entity Tag	No
<b>if-none-match</b>	Gets only if object ETag matches ETag.	Entity Tag	No

## Get Object ACL

### Syntax

```
1. GET /{bucket}/{object}?acl HTTP/1.1
```

### Response Entities

Name	Type	Description
<code>AccessControlPolicy</code>	Container	A container for the response.
<code>AccessControlList</code>	Container	A container for the ACL information.
<code>Owner</code>	Container	A container for the object owner's <code>ID</code> and <code>DisplayName</code> .
<code>ID</code>	String	The object owner's ID.
<code>DisplayName</code>	String	The object owner's display name.
<code>Grant</code>	Container	A container for <code>Grantee</code> and <code>Permission</code> .
<code>Grantee</code>	Container	A container for the <code>DisplayName</code> and <code>ID</code> of the user receiving a grant of permission.
<code>Permission</code>	String	The permission given to the <code>Grantee</code> object.

## Set Object ACL

### Syntax

```
1. PUT /{bucket}/{object}?acl
```

## Request Entities

Name	Type	Description
AccessControlPolicy	Container	A container for the response.
AccessControlList	Container	A container for the ACL information.
Owner	Container	A container for the object owner's ID and DisplayName .
ID	String	The object owner's ID.
DisplayName	String	The object owner's display name.
Grant	Container	A container for Grantee and Permission .
Grantee	Container	A container for the DisplayName and ID of the user receiving a grant of permission.
Permission	String	The permission given to the Grantee object.

# Initiate Multi-part Upload

Initiate a multi-part upload process.

## Syntax

```
1. POST /{bucket}/{object}?uploads
```

## Request Headers

Name	Description	Valid Values	Required
<b>content-md5</b>	A base64 encoded MD-5 hash of the message.	A string. No defaults or constraints.	No
<b>content-type</b>	A standard MIME type.	Any MIME type. Default: <code>binary/octet-stream</code>	No
<b>x-amz-meta-&lt;...&gt;</b>	User metadata. Stored with the object.	A string up to 8kb. No defaults.	No
<b>x-amz-acl</b>	A canned ACL.	<code>private</code> , <code>public-read</code> , <code>public-read-write</code> , <code>authenticated-read</code>	No

## Response Entities

Name	Type	Description
<code>InitiatedMultipartUploadsResult</code>	Container	A container for the results.
<code>Bucket</code>	String	The bucket that will receive the object contents.
<code>Key</code>	String	The key specified by the <code>key</code> request parameter (if any).
<code>UploadId</code>	String	The ID specified by the <code>upload-id</code> request parameter identifying the multipart upload (if any).

# Multipart Upload Part

## Syntax

```
1. PUT /{bucket}/{object}?partNumber=&uploadId= HTTP/1.1
```

## HTTP Response

The following HTTP response may be returned:

HTTP Status	Status Code	Description
<b>404</b>	NoSuchUpload	Specified upload-id does not match any initiated upload on this object

# List Multipart Upload Parts

## Syntax

```
1. GET /{bucket}/{object}?uploadId=123 HTTP/1.1
```

## Response Entities

Name	Type	Description
<code>ListPartsResult</code>	Container	A container for the results.
<code>Bucket</code>	String	The bucket that will receive the object contents.
<code>Key</code>	String	The key specified by the <code>key</code> request parameter (if any).
<code>UploadId</code>	String	The ID specified by the <code>upload-id</code> request parameter identifying the multipart upload (if any).
<code>Initiator</code>	Container	Contains the <code>ID</code> and <code>DisplayName</code> of the user who initiated the upload.
<code>ID</code>	String	The initiator's ID.
<code>DisplayName</code>	String	The initiator's display name.
<code>Owner</code>	Container	A container for the <code>ID</code> and <code>DisplayName</code> of the user who owns the uploaded object.
<code>StorageClass</code>	String	The method used to store the resulting object. <code>STANDARD</code> or <code>REDUCED_REDUNDANCY</code>
<code>PartNumberMarker</code>	String	The part marker to use in a subsequent request if <code>IsTruncated</code> is <code>true</code> . Precedes the list.
<code>NextPartNumberMarker</code>	String	The next part marker to use in a subsequent request if <code>IsTruncated</code> is <code>true</code> . The end of the list.
<code>MaxParts</code>	Integer	The max parts allowed in the response as specified by the <code>max-parts</code> request parameter.
<code>IsTruncated</code>	Boolean	If <code>true</code> , only a subset of the object's upload contents were returned.

Part	Container	A container for <code>LastModified</code> , <code>PartNumber</code> , <code>ETag</code> and <code>Size</code> elements.
<code>LastModified</code>	Date	Date and time at which the part was uploaded.
<code>PartNumber</code>	Integer	The identification number of the part.
<code>ETag</code>	String	The part's entity tag.
<code>Size</code>	Integer	The size of the uploaded part.

# Complete Multipart Upload

Assembles uploaded parts and creates a new object, thereby completing a multipart upload.

## Syntax

```
1. POST /{bucket}/{object}?uploadId= HTTP/1.1
```

## Request Entities

Name	Type	Description	Required
CompleteMultipartUpload	Container	A container consisting of one or more parts.	Yes
Part	Container	A container for the PartNumber and ETag .	Yes
PartNumber	Integer	The identifier of the part.	Yes
ETag	String	The part's entity tag.	Yes

## Response Entities

Name	Type	Description
CompleteMultipartUploadResult	Container	A container for the response.
Location	URI	The resource identifier (path) of the new object.
Bucket	String	The name of the bucket that contains the new object.
Key	String	The object's key.
ETag	String	The entity tag of the new object.

# Abort Multipart Upload

## Syntax

```
1. DELETE /{bucket}/{object}?uploadId= HTTP/1.1
```

# Append Object

Append data to an object. You must have write permissions on the bucket to perform this operation. It is used to upload files in appending mode. The type of the objects created by the Append Object operation is Appendable Object, and the type of the objects uploaded with the Put Object operation is Normal Object. **Append Object can't be used if bucket versioning is enabled or suspended. Synced object will become normal in multisite, but you can still append to the original object. Compression and encryption features are disabled for Appendable objects.**

## Syntax

```
1. PUT /{bucket}/{object}?append&position= HTTP/1.1
```

## Request Headers

Name	Description	Valid Values	Required
<b>content-md5</b>	A base64 encoded MD-5 hash of the message.	A string. No defaults or constraints.	No
<b>content-type</b>	A standard MIME type.	Anv MTME tvne. Default: <b>binary/octet-stream</b>	No
<b>x-amz-meta-&lt;...&gt;</b>	User metadata. Stored with the object.	A string up to 8kb. No defaults.	No
<b>x-amz-acl</b>	A canned ACL.	<b>private , public-read , public-read-write , authenticated-read</b>	No

## Response Headers

Name	Description
<b>x-rgw-next-append-position</b>	Next position to append object

## HTTP Response

The following HTTP response may be returned:

HTTP Status	Status Code	Description
<b>409</b>	PositionNotEqualToLength	Specified position does not match object length
<b>409</b>	ObjectNotAppendable	Specified object can not be appended
<b>409</b>	InvalidBucketstate	Bucket versioning is enabled or suspended

## Put Object Retention

Places an Object Retention configuration on an object.

### Syntax

```
1. PUT /{bucket}/{object}?retention&versionId= HTTP/1.1
```

## Request Entities

Name	Type	Description	Required
<b>Retention</b>	Container	A container for the request.	Yes
<b>Mode</b>	String	Retention mode for the specified object. Valid Values: GOVERNANCE/COMPLIANCE   Yes	
<b>RetainUntilDate</b>	Timestamp	Retention date. Format: 2020-01-05T00:00:00.000Z   Yes	

## Get Object Retention

Gets an Object Retention configuration on an object.

### Syntax

```
1. GET /{bucket}/{object}?retention&versionId= HTTP/1.1
```

## Response Entities

Name	Type	Description	Required
Retention	Container	A container for the request.	Yes
Mode	String	Retention mode for the specified object. Valid Values: GOVERNANCE/COMPLIANCE   Yes	
RetainUntilDate	Timestamp	Retention date. Format: 2020-01-05T00:00:00.000Z   Yes	

## Put Object Legal Hold

Applies a Legal Hold configuration to the specified object.

### Syntax

```
1. PUT /{bucket}/{object}?legal-hold&versionId= HTTP/1.1
```

## Request Entities

Name	Type	Description	Required
LegalHold	Container	A container for the request.	Yes
Status	String	Indicates whether the specified object has a Legal Hold in place. Valid Values: ON/OFF	Yes

## Get Object Legal Hold

Gets an object's current Legal Hold status.

### Syntax

```
1. GET /{bucket}/{object}?legal-hold&versionId= HTTP/1.1
```

## Response Entities

Name	Type	Description	Required
LegalHold	Container	A container for the request.	Yes

Status	String	Indicates whether the specified object has a Legal Hold in place. Valid Values: ON/OFF	Yes
--------	--------	----------------------------------------------------------------------------------------	-----

# C++ S3 Examples

## Setup

The following contains includes and globals that will be used in later examples:

```
1. #include "libs3.h"
2. #include <stdlib.h>
3. #include <iostream>
4. #include <fstream>
5.
6. const char access_key[] = "ACCESS_KEY";
7. const char secret_key[] = "SECRET_KEY";
8. const char host[] = "HOST";
9. const char sample_bucket[] = "sample_bucket";
10. const char sample_key[] = "hello.txt";
11. const char sample_file[] = "resource/hello.txt";
12. const char *security_token = NULL;
13. const char *auth_region = NULL;
14.
15. S3BucketContext bucketContext =
16. {
17.     host,
18.     sample_bucket,
19.     S3ProtocolHTTP,
20.     S3UriStylePath,
21.     access_key,
22.     secret_key,
23.     security_token,
24.     auth_region
25. };
26.
27. S3Status responsePropertiesCallback(
28.         const S3ResponseProperties *properties,
29.         void *callbackData)
30. {
31.     return S3StatusOK;
32. }
33.
34. static void responseCompleteCallback(
35.         S3Status status,
36.         const S3ErrorDetails *error,
37.         void *callbackData)
38. {
39.     return;
40. }
41.
42. S3ResponseHandler responseHandler =
43. {
```

```

44.     &responsePropertiesCallback,
45.     &responseCompleteCallback
46. };

```

## Creating (and Closing) a Connection

This creates a connection so that you can interact with the server.

```

1. S3_initialize("s3", S3_INIT_ALL, host);
2. // Do stuff...
3. S3_deinitialize();

```

## Listing Owned Buckets

This gets a list of Buckets that you own. This also prints out the bucket name, owner ID, and display name for each bucket.

```

1. static S3Status listServiceCallback(
2.         const char *ownerId,
3.         const char *ownerDisplayName,
4.         const char *bucketName,
5.         int64_t creationDate, void *callbackData)
6. {
7.     bool *header_printed = (bool*) callbackData;
8.     if (!*header_printed) {
9.         *header_printed = true;
10.        printf("%-22s", "      Bucket");
11.        printf(" %-20s %-12s", "      Owner ID", "Display Name");
12.        printf("\n");
13.        printf("-----");
14.        printf(" -----" " -----");
15.        printf("\n");
16.    }
17.
18.    printf("%-22s", bucketName);
19.    printf(" %-20s %-12s", ownerId ? ownerId : "", ownerDisplayName ? ownerDisplayName : "");
20.    printf("\n");
21.
22.    return S3StatusOK;
23. }
24.
25. S3ListServiceHandler listServiceHandler =
26. {
27.     responseHandler,
28.     &listServiceCallback
29. };
30. bool header_printed = false;
31. S3_list_service(S3ProtocolHTTP, access_key, secret_key, security_token, host,
32.                 auth_region, NULL, 0, &listServiceHandler, &header_printed);

```

# Creating a Bucket

This creates a new bucket.

```
1. S3_create_bucket(S3ProtocolHTTP, access_key, secret_key, NULL, host, sample_bucket, S3CannedAclPrivate, NULL,
1.     NULL, &responseHandler, NULL);
```

# Listing a Bucket's Content

This gets a list of objects in the bucket. This also prints out each object's name, the file size, and last modified date.

```
1. static S3Status listBucketCallback(
2.         int isTruncated,
3.         const char *nextMarker,
4.         int contentsCount,
5.         const S3ListBucketContent *contents,
6.         int commonPrefixesCount,
7.         const char **commonPrefixes,
8.         void *callbackData)
9. {
10.     printf("%-22s", "      Object Name");
11.     printf(" % -5s % -20s", "Size", " Last Modified");
12.     printf("\n");
13.     printf("-----");
14.     printf(" -----" " -----");
15.     printf("\n");
16.
17.     for (int i = 0; i < contentsCount; i++) {
18.         char timebuf[256];
19.         char sizebuf[16];
20.         const S3ListBucketContent *content = &(contents[i]);
21.         time_t t = (time_t) content->lastModified;
22.
23.         strftime(timebuf, sizeof(timebuf), "%Y-%m-%dT%H:%M:%SZ", gmtime(&t));
24.         sprintf(sizebuf, "%5llu", (unsigned long long) content->size);
25.         printf("%-22s %s %s\n", content->key, sizebuf, timebuf);
26.     }
27.
28.     return S3StatusOK;
29. }
30.
31. S3ListBucketHandler listBucketHandler =
32. {
33.     responseHandler,
34.     &listBucketCallback
35. };
36. S3_list_bucket(&bucketContext, NULL, NULL, NULL, 0, NULL, 0, &listBucketHandler, NULL);
```

The output will look something like this:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
S3_delete_bucket(S3ProtocolHTTP, S3UriStylePath, access_key, secret_key, 0, host, sample_bucket, NULL, NULL,
1. 0, &responseHandler, NULL);
```

## Creating an Object (from a file)

This creates a file `hello.txt`.

```
1. #include <sys/stat.h>
2. typedef struct put_object_callback_data
3. {
4.     FILE *infile;
5.     uint64_t contentLength;
6. } put_object_callback_data;
7.
8.
9. static int putObjectDataCallback(int bufferSize, char *buffer, void *callbackData)
10. {
11.     put_object_callback_data *data = (put_object_callback_data *) callbackData;
12.
13.     int ret = 0;
14.
15.     if (data->contentLength) {
16.         int toRead = ((data->contentLength > (unsigned) bufferSize) ? (unsigned) bufferSize : data-
17. >contentLength);
18.         ret = fread(buffer, 1, toRead, data->infile);
19.         data->contentLength -= ret;
20.     }
21.     return ret;
22. }
23. put_object_callback_data data;
24. struct stat statbuf;
25. if (stat(sample_file, &statbuf) == -1) {
26.     fprintf(stderr, "\nERROR: Failed to stat file %s: ", sample_file);
27.     perror(0);
28.     exit(-1);
29. }
30.
```

```

31. int contentLength = statbuf.st_size;
32. data.contentLength = contentLength;
33.
34. if (!(data.infile = fopen(sample_file, "r")) ) {
35.     fprintf(stderr, "\nERROR: Failed to open input file %s: ", sample_file);
36.     perror(0);
37.     exit(-1);
38. }
39.
40. S3PutObjectHandler putObjectHandler =
41. {
42.     responseHandler,
43.     &putObjectDataCallback
44. };
45.
46. S3_put_object(&bucketContext, sample_key, contentLength, NULL, NULL, 0, &putObjectHandler, &data);
47. fclose(data.infile);

```

## Download an Object (to a file)

This downloads a file and prints the contents.

```

1. static S3Status getObjectDataCallback(int bufferSize, const char *buffer, void *callbackData)
2. {
3.     FILE *outfile = (FILE *) callbackData;
4.     size_t wrote = fwrite(buffer, 1, bufferSize, outfile);
5.     return ((wrote < (size_t) bufferSize) ? S3StatusAbortedByCallback : S3StatusOK);
6. }
7.
8. S3GetObjectHandler getObjectHandler =
9. {
10.     responseHandler,
11.     &getObjectDataCallback
12. };
13. FILE *outfile = stdout;
14. S3_get_object(&bucketContext, sample_key, NULL, 0, 0, NULL, 0, &getObjectHandler, outfile);

```

## Delete an Object

This deletes an object.

```

1. S3ResponseHandler deleteResponseHandler =
2. {
3.     0,
4.     &responseCompleteCallback
5. };
6. S3_delete_object(&bucketContext, sample_key, 0, 0, &deleteResponseHandler, 0);

```

# Change an Object's ACL

This changes an object's ACL to grant full control to another user.

```

1. #include <string.h>
2. char ownerId[] = "owner";
3. char ownerDisplayName[] = "owner";
4. char granteeId[] = "grantee";
5. char granteeDisplayName[] = "grantee";
6.
7. S3AclGrant grants[] = {
8.     {
9.         S3GranteeTypeCanonicalUser,
10.        {{}},
11.        S3PermissionFullControl
12.    },
13.    {
14.        S3GranteeTypeCanonicalUser,
15.        {{}},
16.        S3PermissionReadACP
17.    },
18.    {
19.        S3GranteeTypeAllUsers,
20.        {{}},
21.        S3PermissionRead
22.    }
23. };
24.
25. strncpy(grants[0].grantee.canonicalUser.id, ownerId, S3_MAX_GRANTEE_USER_ID_SIZE);
26. strncpy(grants[0].grantee.canonicalUser.displayName, ownerDisplayName, S3_MAX_GRANTEE_DISPLAY_NAME_SIZE);
27.
28. strncpy(grants[1].grantee.canonicalUser.id, granteeId, S3_MAX_GRANTEE_USER_ID_SIZE);
29. strncpy(grants[1].grantee.canonicalUser.displayName, granteeDisplayName, S3_MAX_GRANTEE_DISPLAY_NAME_SIZE);
30.
31. S3_set_acl(&bucketContext, sample_key, ownerId, ownerDisplayName, 3, grants, 0, &responseHandler, 0);

```

# Generate Object Download URL (signed)

This generates a signed download URL that will be valid for 5 minutes.

```

1. #include <time.h>
2. char buffer[S3_MAX_AUTHENTICATED_QUERY_STRING_SIZE];
3. int64_t expires = time(NULL) + 60 * 5; // Current time + 5 minutes
4.
5. S3_generate_authenticated_query_string(buffer, &bucketContext, sample_key, expires, NULL, "GET");

```

# C# S3 Examples

## Creating a Connection

This creates a connection so that you can interact with the server.

```
1. using System;
2. using Amazon;
3. using Amazon.S3;
4. using Amazon.S3.Model;
5.
6. string accessKey = "put your access key here!";
7. string secretKey = "put your secret key here!";
8.
9. AmazonS3Config config = new AmazonS3Config();
10. config.ServiceURL = "objects.dreamhost.com";
11.
12. AmazonS3Client s3Client = new AmazonS3Client(
13.     accessKey,
14.     secretKey,
15.     config
16. );
```

## Listing Owned Buckets

This gets a list of Buckets that you own. This also prints out the bucket name and creation date of each bucket.

```
1. ListBucketsResponse response = client.ListBuckets();
2. foreach (S3Bucket b in response.Buckets)
3. {
4.     Console.WriteLine("{0}\t{1}", b.BucketName, b.CreationDate);
5. }
```

The output will look something like this:

```
1. mahbucket1 2011-04-21T18:05:39.000Z
2. mahbucket2 2011-04-21T18:05:48.000Z
3. mahbucket3 2011-04-21T18:07:18.000Z
```

## Creating a Bucket

This creates a new bucket called `my-new-bucket`

```

1. PutBucketRequest request = new PutBucketRequest();
2. request.BucketName = "my-new-bucket";
3. client.PutBucket(request);

```

## Listing a Bucket's Content

This gets a list of objects in the bucket. This also prints out each object's name, the file size, and last modified date.

```

1. ListObjectsRequest request = new ListObjectsRequest();
2. request.BucketName = "my-new-bucket";
3. ListObjectsResponse response = client.ListObjects(request);
4. foreach (S3Object o in response.S3Objects)
5. {
6.     Console.WriteLine("{0}\t{1}\t{2}", o.Key, o.Size, o.LastModified);
7. }

```

The output will look something like this:

```

1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z

```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```

1. DeleteBucketRequest request = new DeleteBucketRequest();
2. request.BucketName = "my-new-bucket";
3. client.DeleteBucket(request);

```

## Forced Delete for Non-empty Buckets

### Attention

not available

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```

1. PutObjectRequest request = new PutObjectRequest();
2. request.BucketName = "my-new-bucket";

```

```

3. request.Key      = "hello.txt";
4. request.ContentType = "text/plain";
5. request.ContentBody = "Hello World!";
6. client.PutObject(request);

```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable, and `secret_plans.txt` to be private.

```

1. PutACLRequest request = new PutACLRequest();
2. request.BucketName = "my-new-bucket";
3. request.Key      = "hello.txt";
4. request.CannedACL = S3CannedACL.PublicRead;
5. client.PutACL(request);
6.
7. PutACLRequest request2 = new PutACLRequest();
8. request2.BucketName = "my-new-bucket";
9. request2.Key      = "secret_plans.txt";
10. request2.CannedACL = S3CannedACL.Private;
11. client.PutACL(request2);

```

## Download an Object (to a file)

This downloads the object `perl_poetry.pdf` and saves it in `C:\Users\larry\Documents`

```

1. GetObjectRequest request = new GetObjectRequest();
2. request.BucketName = "my-new-bucket";
3. request.Key      = "perl_poetry.pdf";
4. GetObjectResponse response = client.GetObject(request);
5. response.WriteResponseStreamToFile("C:\\\\Users\\\\larry\\\\Documents\\\\perl_poetry.pdf");

```

## Delete an Object

This deletes the object `goodbye.txt`

```

1. DeleteObjectRequest request = new DeleteObjectRequest();
2. request.BucketName = "my-new-bucket";
3. request.Key      = "goodbye.txt";
4. client.DeleteObject(request);

```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

#### Note

The C# S3 Library does not have a method for generating unsigned URLs, so the following example only shows generating signed URLs.

```
1. GetPreSignedUrlRequest request = new GetPreSignedUrlRequest();
2. request.BucketName = "my-bucket-name";
3. request.Key      = "secret_plans.txt";
4. request.Expires = DateTime.Now.AddHours(1);
5. request.Protocol = Protocol.HTTP;
6. string url = client.GetPreSignedURL(request);
7. Console.WriteLine(url);
```

The output of this will look something like:

```
http://objects.dreamhost.com/my-bucket-name/secret_plans.txt?
1. Signature=XXXXXXXXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX
```

# Java S3 Examples

## Setup

The following examples may require some or all of the following java classes to be imported:

```
1. import java.io.ByteArrayInputStream;
2. import java.io.File;
3. import java.util.List;
4. import com.amazonaws.auth.AWS Credentials;
5. import com.amazonaws.auth.BasicAWSCredentials;
6. import com.amazonaws.util.StringUtils;
7. import com.amazonaws.services.s3.AmazonS3;
8. import com.amazonaws.services.s3.AmazonS3Client;
9. import com.amazonaws.services.s3.model.Bucket;
10. import com.amazonaws.services.s3.model.CannedAccessControlList;
11. import com.amazonaws.services.s3.model.GeneratePresignedUrlRequest;
12. import com.amazonaws.services.s3.model.GetObjectRequest;
13. import com.amazonaws.services.s3.model.ObjectListing;
14. import com.amazonaws.services.s3.model.ObjectMetadata;
15. import com.amazonaws.services.s3.model.S3ObjectSummary;
```

If you are just testing the Ceph Object Storage services, consider using HTTP protocol instead of HTTPS protocol.

First, import the `ClientConfiguration` and `Protocol` classes.

```
1. import com.amazonaws.ClientConfiguration;
2. import com.amazonaws.Protocol;
```

Then, define the client configuration, and add the client configuration as an argument for the S3 client.

```
1. AWS Credentials credentials = new BasicAWSCredentials(accessKey, secretKey);
2.
3. ClientConfiguration clientConfig = new ClientConfiguration();
4. clientConfig.setProtocol(Protocol.HTTP);
5.
6. AmazonS3 conn = new AmazonS3Client(credentials, clientConfig);
7. conn.setEndpoint("endpoint.com");
```

## Creating a Connection

This creates a connection so that you can interact with the server.

```

1. String accessKey = "insert your access key here!";
2. String secretKey = "insert your secret key here!";
3.
4. AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
5. AmazonS3 conn = new AmazonS3Client(credentials);
6. conn.setEndpoint("objects.dreamhost.com");

```

## Listing Owned Buckets

This gets a list of Buckets that you own. This also prints out the bucket name and creation date of each bucket.

```

1. List<Bucket> buckets = conn.listBuckets();
2. for (Bucket bucket : buckets) {
3.     System.out.println(bucket.getName() + "\t" +
4.         StringUtils.fromDate(bucket.getCreationDate()));
5. }

```

The output will look something like this:

```

1. mahbucket1  2011-04-21T18:05:39.000Z
2. mahbucket2  2011-04-21T18:05:48.000Z
3. mahbucket3  2011-04-21T18:07:18.000Z

```

## Creating a Bucket

This creates a new bucket called `my-new-bucket`

```
1. Bucket bucket = conn.createBucket("my-new-bucket");
```

## Listing a Bucket's Content

This gets a list of objects in the bucket. This also prints out each object's name, the file size, and last modified date.

```

1. ObjectListing objects = conn.listObjects(bucket.getName());
2. do {
3.     for (S3ObjectSummary objectSummary : objects.getObjectSummaries()) {
4.         System.out.println(objectSummary.getKey() + "\t" +
5.             objectSummary.getSize() + "\t" +
6.             StringUtils.fromDate(objectSummary.getLastModified()));
7.     }
8.     objects = conn.listNextBatchOfObjects(objects);
9. } while (objects.isTruncated());

```

The output will look something like this:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
1. conn.deleteBucket(bucket.getName());
```

## Forced Delete for Non-empty Buckets

### Attention

not available

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```
1. ByteArrayInputStream input = new ByteArrayInputStream("Hello World!".getBytes());
2. conn.putObject(bucket.getName(), "hello.txt", input, new ObjectMetadata());
```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable, and `secret_plans.txt` to be private.

```
1. conn.setObjectAcl(bucket.getName(), "hello.txt", CannedAccessControlList.PublicRead);
2. conn.setObjectAcl(bucket.getName(), "secret_plans.txt", CannedAccessControlList.Private);
```

## Download an Object (to a file)

This downloads the object `perl_poetry.pdf` and saves it in `/home/larry/documents`

```
1. conn.getObject(
2.     new GetObjectRequest(bucket.getName(), "perl_poetry.pdf"),
3.     new File("/home/larry/documents/perl_poetry.pdf")
4. );
```

# Delete an Object

This deletes the object `goodbye.txt`

```
1. conn.deleteObject(bucket.getName(), "goodbye.txt");
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

### Note

The java library does not have a method for generating unsigned URLs, so the example below just generates a signed URL.

```
1. GeneratePresignedUrlRequest request = new GeneratePresignedUrlRequest(bucket.getName(), "secret_plans.txt");
2. System.out.println(conn.generatePresignedUrl(request));
```

The output will look something like this:

```
https://my-bucket-name.objects.dreamhost.com/secret_plans.txt?
1. Signature=XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX
```

# Perl S3 Examples

## Creating a Connection

This creates a connection so that you can interact with the server.

```

1. use Amazon::S3;
2. my $access_key = 'put your access key here!';
3. my $secret_key = 'put your secret key here!';
4.
5. my $conn = Amazon::S3->new({
6.     aws_access_key_id      => $access_key,
7.     aws_secret_access_key => $secret_key,
8.     host                  => 'objects.dreamhost.com',
9.     secure                => 1,
10.    retry                 => 1,
11. });

```

## Listing Owned Buckets

This gets a list of `Amazon::S3::Bucket` objects that you own. We'll also print out the bucket name and creation date of each bucket.

```

1. my @buckets = @{$conn->buckets->{buckets} || []};
2. foreach my $bucket (@buckets) {
3.     print $bucket->bucket . "\t" . $bucket->creation_date . "\n";
4. }

```

The output will look something like this:

```

1. mahbuckat1  2011-04-21T18:05:39.000Z
2. mahbuckat2  2011-04-21T18:05:48.000Z
3. mahbuckat3  2011-04-21T18:07:18.000Z

```

## Creating a Bucket

This creates a new bucket called `my-new-bucket`

```
1. my $bucket = $conn->add_bucket({ bucket => 'my-new-bucket' });
```

## Listing a Bucket's Content

This gets a list of hashes with info about each object in the bucket. We'll also print out each object's name, the file size, and last modified date.

```
1. my @keys = @{$bucket->list_all->{keys} || []};
2. foreach my $key (@keys) {
3.     print "$key->{key}\t$key->{size}\t$key->{last_modified}\n";
4. }
```

The output will look something like this:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
1. $conn->delete_bucket($bucket);
```

## Forced Delete for Non-empty Buckets

### Attention

not available in the `Amazon::S3` perl module

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```
1. $bucket->add_key(
2.     'hello.txt', 'Hello World!',
3.     { content_type => 'text/plain' },
4. );
```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable and `secret_plans.txt` to be private.

```
1. $bucket->set_acl({
2.     key      => 'hello.txt',
3.     acl_short => 'public-read',
```

```

4. });
5. $bucket->set_acl({
6.     key      => 'secret_plans.txt',
7.     acl_short => 'private',
8. });

```

## Download an Object (to a file)

This downloads the object `perl_poetry.pdf` and saves it in `/home/larry/documents/`

```

1. $bucket->get_key_filename('perl_poetry.pdf', undef,
2.                            '/home/larry/documents/perl_poetry.pdf');

```

## Delete an Object

This deletes the object `goodbye.txt`

```
1. $bucket->delete_key('goodbye.txt');
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. Then this generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

### Note

The `Amazon::S3` module does not have a way to generate download URLs, so we are going to be using another module instead. Unfortunately, most modules for generating these URLs assume that you are using Amazon, so we have had to go with using a more obscure module, `Muck::FS::S3`. This should be the same as Amazon's sample S3 perl module, but this sample module is not in CPAN. So, you can either use CPAN to install `Muck::FS::S3`, or install Amazon's sample S3 module manually. If you go the manual route, you can remove `Muck::FS::` from the example below.

```

1. use Muck::FS::S3::QueryStringAuthGenerator;
2. my $generator = Muck::FS::S3::QueryStringAuthGenerator->new(
3.     $access_key,
4.     $secret_key,
5.     0, # 0 means use 'http'. set this to 1 for 'https'
6.     'objects.dreamhost.com',
7. );

```

```
8.  
9. my $hello_url = $generator->make_bare_url($bucket->bucket, 'hello.txt');  
10. print $hello_url . "\n";  
11.  
12. $generator->expires_in(3600); # 1 hour = 3600 seconds  
13. my $plans_url = $generator->get($bucket->bucket, 'secret_plans.txt');  
14. print $plans_url . "\n";
```

The output will look something like this:

```
1. http://objects.dreamhost.com:80/my-bucket-name/hello.txt  
http://objects.dreamhost.com:80/my-bucket-name/secret_plans.txt?  
2. Signature=XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX
```

# PHP S3 Examples

## Installing AWS PHP SDK

This installs AWS PHP SDK using composer (see [here](#) how to install composer).

```
1. $ composer install aws/aws-sdk-php
```

## Creating a Connection

This creates a connection so that you can interact with the server.

### Note

The client initialization requires a region so we use `'us-east-1'`.

```
1. <?php
2.
3. use Aws\S3\S3Client;
4.
5. define('AWS_KEY', 'place access key here');
6. define('AWS_SECRET_KEY', 'place secret key here');
7. $ENDPOINT = 'http://objects.dreamhost.com';
8.
9. // require the amazon sdk from your composer vendor dir
10. require __DIR__.'/vendor/autoload.php';
11.
12. // Instantiate the S3 class and point it at the desired host
13. $client = new S3Client([
14.     'region' => '',
15.     'version' => '2006-03-01',
16.     'endpoint' => $ENDPOINT,
17.     'credentials' => [
18.         'key' => AWS_KEY,
19.         'secret' => AWS_SECRET_KEY
20.     ],
21.     // Set the S3 class to use objects.dreamhost.com/bucket
22.     // instead of bucket.objects.dreamhost.com
23.     'use_path_style_endpoint' => true
24. ]);
```

## Listing Owned Buckets

This gets a `AWS\Result` instance that is more convenient to visit using array access way. This also prints out the bucket name and creation date of each bucket.

```

1. <?php
2. $listResponse = $client->listBuckets();
3. $buckets = $listResponse['Buckets'];
4. foreach ($buckets as $bucket) {
5.     echo $bucket['Name'] . "\t" . $bucket['CreationDate'] . "\n";
6. }

```

The output will look something like this:

```

1. mahbuckat1 2011-04-21T18:05:39.000Z
2. mahbuckat2 2011-04-21T18:05:48.000Z
3. mahbuckat3 2011-04-21T18:07:18.000Z

```

## Creating a Bucket

This creates a new bucket called `my-new-bucket` and returns a `AWS\Result` object.

```

1. <?php
2. $client->createBucket(['Bucket' => 'my-new-bucket']);

```

## List a Bucket's Content

This gets a `AWS\Result` instance that is more convenient to visit using array access way. This then prints out each object's name, the file size, and last modified date.

```

1. <?php
2. $objectsListResponse = $client->listObjects(['Bucket' => $bucketname]);
3. $objects = $objectsListResponse['Contents'] ?? [];
4. foreach ($objects as $object) {
5.     echo $object['Key'] . "\t" . $object['Size'] . "\t" . $object['LastModified'] . "\n";
6. }

```

### Note

If there are more than 1000 objects in this bucket, you need to check `$objectsListResponse['isTruncated']` and run again with the name of the last key listed. Keep doing this until `isTruncated` is not true.

The output will look something like this if the bucket has some files:

```

1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z

```

## Deleting a Bucket

This deletes the bucket called `my-old-bucket` and returns a `AWS\Result` object

## Note

The Bucket must be empty! Otherwise it won't work!

```
1. <?php
2. $client->deleteBucket(['Bucket' => 'my-old-bucket']);
```

## Creating an Object

This creates an object `hello.txt` with the string `"Hello World!"`

```
1. <?php
2. $client->putObject([
3.     'Bucket' => 'my-bucket-name',
4.     'Key' => 'hello.txt',
5.     'Body' => "Hello World!"
6. ]);
```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable and `secret_plans.txt` to be private.

```
1. <?php
2. $client->putObjectAcl([
3.     'Bucket' => 'my-bucket-name',
4.     'Key' => 'hello.txt',
5.     'ACL' => 'public-read'
6. ]);
7. $client->putObjectAcl([
8.     'Bucket' => 'my-bucket-name',
9.     'Key' => 'secret_plans.txt',
10.    'ACL' => 'private'
11.]);
```

## Delete an Object

This deletes the object `goodbye.txt`

```
1. <?php
2. $client->deleteObject(['Bucket' => 'my-bucket-name', 'Key' => 'goodbye.txt']);
```

## Download an Object (to a file)

This downloads the object `poetry.pdf` and saves it in `/home/larry/documents/`

```
1. <?php
2. $object = $client->getObject(['Bucket' => 'my-bucket-name', 'Key' => 'poetry.pdf']);
3. file_put_contents('/home/larry/documents/poetry.pdf', $object['Body']->getContents());
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

```
1. <?php
2. $hello_url = $client->getObjectUrl('my-bucket-name', 'hello.txt');
3. echo $hello_url."\n";
4.
5. $secret_plans_cmd = $client->getCommand('GetObject', ['Bucket' => 'my-bucket-name', 'Key' =>
6.   'secret_plans.txt']);
7. $request = $client->createPresignedRequest($secret_plans_cmd, '+1 hour');
8. echo $request->getUri()."\n";
```

The output of this will look something like:

```
1. http://objects.dreamhost.com/my-bucket-name/hello.txt
http://objects.dreamhost.com/my-bucket-name/secret_plans.txt?X-Amz-Content-Sha256=UNSIGNED-PAYLOAD&X-Amz-
Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=sandboxAccessKey%2F20190116%2F%2Fs3%2Faws4_request&X-Amz-
Date=20190116T125520Z&X-Amz-SignedHeaders=host&X-Amz-Expires=3600&X-Amz-
2. Signature=61921f07c73d7695e47a2192cf55ae030f34c44c512b2160bb5a936b2b48d923
```

# Python S3 Examples

## Creating a Connection

This creates a connection so that you can interact with the server.

```

1. import boto
2. import boto.s3.connection
3. access_key = 'put your access key here!'
4. secret_key = 'put your secret key here!'
5.
6. conn = boto.connect_s3(
7.     aws_access_key_id = access_key,
8.     aws_secret_access_key = secret_key,
9.     host = 'objects.dreamhost.com',
10.    #is_secure=False,           # uncomment if you are not using ssl
11.    calling_format = boto.s3.connection.OrdinaryCallingFormat(),
12. )

```

## Listing Owned Buckets

This gets a list of Buckets that you own. This also prints out the bucket name and creation date of each bucket.

```

1. for bucket in conn.get_all_buckets():
2.     print "{name}\t{created}".format(
3.         name = bucket.name,
4.         created = bucket.creation_date,
5.     )

```

The output will look something like this:

```

1. mahbucket1  2011-04-21T18:05:39.000Z
2. mahbucket2  2011-04-21T18:05:48.000Z
3. mahbucket3  2011-04-21T18:07:18.000Z

```

## Creating a Bucket

This creates a new bucket called `my-new-bucket`

```
1. bucket = conn.create_bucket('my-new-bucket')
```

## Listing a Bucket's Content

This gets a list of objects in the bucket. This also prints out each object's name, the file size, and last modified date.

```

1. for key in bucket.list():
2.     print "{name}\t{size}\t{modified}".format(
3.         name = key.name,
4.         size = key.size,
5.         modified = key.last_modified,
6.     )

```

The output will look something like this:

```

1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z

```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
1. conn.delete_bucket(bucket.name)
```

## Forced Delete for Non-empty Buckets

### Attention

not available in python

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```

1. key = bucket.new_key('hello.txt')
2. key.set_contents_from_string('Hello World!')

```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable, and `secret_plans.txt` to be private.

```

1. hello_key = bucket.get_key('hello.txt')
2. hello_key.set_canned_acl('public-read')
3. plans_key = bucket.get_key('secret_plans.txt')
4. plans_key.set_canned_acl('private')

```

## Download an Object (to a file)

This downloads the object `perl_poetry.pdf` and saves it in `/home/larry/documents/`

```

1. key = bucket.get_key('perl_poetry.pdf')
2. key.get_contents_to_filename('/home/larry/documents/perl_poetry.pdf')

```

## Delete an Object

This deletes the object `goodbye.txt`

```
1. bucket.delete_key('goodbye.txt')
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

```

1. hello_key = bucket.get_key('hello.txt')
2. hello_url = hello_key.generate_url(0, query_auth=False, force_http=True)
3. print hello_url
4.
5. plans_key = bucket.get_key('secret_plans.txt')
6. plans_url = plans_key.generate_url(3600, query_auth=True, force_http=True)
7. print plans_url

```

The output of this will look something like:

```

1. http://objects.dreamhost.com/my-bucket-name/hello.txt
   http://objects.dreamhost.com/my-bucket-name/secret_plans.txt?
2. Signature=XXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX

```

## Using S3 API Extensions

To use the boto3 client to tests the RadosGW extensions to the S3 API, the [extensions file](#) should be placed under: `~/.aws/models/s3/2006-03-01/` directory. For example, unordered list of objects could be fetched using:

```
1. print conn.list_objects(Bucket='my-new-bucket', AllowUnordered=True)
```

Without the extensions file, in the above example, boto3 would complain that the `AllowUnordered` argument is invalid.

# Ruby AWS::SDK Examples (aws-sdk gem ~>2)

## Settings

You can setup the connection on global way:

```
1. Aws.config.update(  
2.     endpoint: 'https://objects.dreamhost.com.',  
3.     access_key_id: 'my-access-key',  
4.     secret_access_key: 'my-secret-key',  
5.     force_path_style: true,  
6.     region: 'us-east-1'  
7. )
```

and instantiate a client object:

```
1. s3_client = Aws::S3::Client.new
```

## Listing Owned Buckets

This gets a list of buckets that you own. This also prints out the bucket name and creation date of each bucket.

```
1. s3_client.list_buckets.buckets.each do |bucket|  
2.     puts "#{bucket.name}\t#{bucket.creation_date}"  
3. end
```

The output will look something like this:

```
1. mahbuckat1  2011-04-21T18:05:39.000Z  
2. mahbuckat2  2011-04-21T18:05:48.000Z  
3. mahbuckat3  2011-04-21T18:07:18.000Z
```

## Creating a Bucket

This creates a new bucket called `my-new-bucket`

```
1. s3_client.create_bucket(bucket: 'my-new-bucket')
```

If you want a private bucket:

acl option accepts: # private, public-read, public-read-write, authenticated-read

```
1. s3_client.create_bucket(bucket: 'my-new-bucket', acl: 'private')
```

## Listing a Bucket's Content

This gets a list of hashes with the contents of each object. This also prints out each object's name, the file size, and last modified date.

```
1. s3_client.get_objects(bucket: 'my-new-bucket').contents.each do |object|
2.   puts "#{object.key}\t#{object.size}\t#{object.last_modified}"
3. end
```

The output will look something like this if the bucket has some files:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
1. s3_client.delete_bucket(bucket: 'my-new-bucket')
```

## Forced Delete for Non-empty Buckets

First, you need to clear the bucket:

```
1. Aws::S3::Bucket.new('my-new-bucket', client: s3_client).clear!
```

after, you can destroy the bucket

```
1. s3_client.delete_bucket(bucket: 'my-new-bucket')
```

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```
1. s3_client.put_object(
2.   key: 'hello.txt',
3.   body: 'Hello World!',
4.   bucket: 'my-new-bucket',
5.   content_type: 'text/plain'
```

6. )

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable, and `secret_plans.txt` to be private.

```
1. s3_client.put_object_acl(bucket: 'my-new-bucket', key: 'hello.txt', acl: 'public-read')
2.
3. s3_client.put_object_acl(bucket: 'my-new-bucket', key: 'private.txt', acl: 'private')
```

## Download an Object (to a file)

This downloads the object `poetry.pdf` and saves it in `/home/larry/documents/`

```
s3_client.get_object(bucket: 'my-new-bucket', key: 'poetry.pdf', response_target:
1. '/home/larry/documents/poetry.pdf')
```

## Delete an Object

This deletes the object `goodbye.txt`

```
1. s3_client.delete_object(key: 'goodbye.txt', bucket: 'my-new-bucket')
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

```
1. puts Aws::S3::Object.new(
2.   key: 'hello.txt',
3.   bucket_name: 'my-new-bucket',
4.   client: s3_client
5. ).public_url
6.
7. puts Aws::S3::Object.new(
8.   key: 'secret_plans.txt',
9.   bucket_name: 'hermes_ceph_gem',
10.  client: s3_client
11. ).presigned_url(:get, expires_in: 60 * 60)
```

The output of this will look something like:

```
1. http://objects.dreamhost.com/my-bucket-name/hello.txt
   http://objects.dreamhost.com/my-bucket-name/secret_plans.txt?
2. Signature=XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX
```

## Ruby AWS::S3 Examples (aws-s3 gem)

### Creating a Connection

This creates a connection so that you can interact with the server.

```
1. AWS::S3::Base.establish_connection(
2.   :server          => 'objects.dreamhost.com',
3.   :use_ssl         => true,
4.   :access_key_id  => 'my-access-key',
5.   :secret_access_key => 'my-secret-key'
6. )
```

### Listing Owned Buckets

This gets a list of `AWS::S3::Bucket` objects that you own. This also prints out the bucket name and creation date of each bucket.

```
1. AWS::S3::Service.buckets.each do |bucket|
2.   puts "#{bucket.name}\t#{bucket.creation_date}"
3. end
```

The output will look something like this:

```
1. mahbucket1  2011-04-21T18:05:39.000Z
2. mahbucket2  2011-04-21T18:05:48.000Z
3. mahbucket3  2011-04-21T18:07:18.000Z
```

### Creating a Bucket

This creates a new bucket called `my-new-bucket`

```
1. AWS::S3::Bucket.create('my-new-bucket')
```

### Listing a Bucket's Content

This gets a list of hashes with the contents of each object. This also prints out each object's name, the file size, and last modified date.

```
1. new_bucket = AWS::S3::Bucket.find('my-new-bucket')
2. new_bucket.each do |object|
3.   puts "#{object.key}\t#{object.size}\t#{object.last_modified}"
4. end
```

The output will look something like this if the bucket has some files:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Deleting a Bucket

### Note

The Bucket must be empty! Otherwise it won't work!

```
1. AWS::S3::Bucket.delete('my-new-bucket')
```

## Forced Delete for Non-empty Buckets

```
1. AWS::S3::Bucket.delete('my-new-bucket', :force => true)
```

## Creating an Object

This creates a file `hello.txt` with the string `"Hello World!"`

```
1. AWS::S3::S3Object.store(
2.   'hello.txt',
3.   'Hello World!',
4.   'my-new-bucket',
5.   :content_type => 'text/plain'
6. )
```

## Change an Object's ACL

This makes the object `hello.txt` to be publicly readable, and `secret_plans.txt` to be private.

```
1. policy = AWS::S3::S3Object.acl('hello.txt', 'my-new-bucket')
2. policy.grants = [ AWS::S3::ACL::Grant.grant(:public_read) ]
```

```

3. AWS::S3::S3Object.acl('hello.txt', 'my-new-bucket', policy)
4.
5. policy = AWS::S3::S3Object.acl('secret_plans.txt', 'my-new-bucket')
6. policy.grants = []
7. AWS::S3::S3Object.acl('secret_plans.txt', 'my-new-bucket', policy)

```

## Download an Object (to a file)

This downloads the object `poetry.pdf` and saves it in `/home/larry/documents/`

```

1. open('/home/larry/documents/poetry.pdf', 'w') do |file|
2.   AWS::S3::S3Object.stream('poetry.pdf', 'my-new-bucket') do |chunk|
3.     file.write(chunk)
4.   end
5. end

```

## Delete an Object

This deletes the object `goodbye.txt`

```
1. AWS::S3::S3Object.delete('goodbye.txt', 'my-new-bucket')
```

## Generate Object Download URLs (signed and unsigned)

This generates an unsigned download URL for `hello.txt`. This works because we made `hello.txt` public by setting the ACL above. This then generates a signed download URL for `secret_plans.txt` that will work for 1 hour. Signed download URLs will work for the time period even if the object is private (when the time period is up, the URL will stop working).

```

1. puts AWS::S3::S3Object.url_for(
2.   'hello.txt',
3.   'my-new-bucket',
4.   :authenticated => false
5. )
6.
7. puts AWS::S3::S3Object.url_for(
8.   'secret_plans.txt',
9.   'my-new-bucket',
10.  :expires_in => 60 * 60
11. )

```

The output of this will look something like:

1. `http://objects.dreamhost.com/my-bucket-name/hello.txt`  
`http://objects.dreamhost.com/my-bucket-name/secret_plans.txt?`
2. `Signature=XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX&Expires=1316027075&AWSAccessKeyId=XXXXXXXXXXXXXXXXXXXX`

# RGW Data caching and CDN

New in version Octopus.

## Contents

- RGW Data caching and CDN

- New APIs

- Auth API

- Cache API

- Using Nginx with RGW

- Appendix

This feature adds to RGW the ability to securely cache objects and offload the workload from the cluster, using Nginx. After an object is accessed the first time it will be stored in the Nginx cache directory. When data is already cached, it need not be fetched from RGW. A permission check will be made against RGW to ensure the requesting user has access. This feature is based on some Nginx modules, `ngx_http_auth_request_module`, <https://github.com/kaltura/nginx-aws-auth-module>, Openresty for Lua capabilities.

Currently, this feature will cache only AWSv4 requests (only s3 requests), caching-in the output of the 1st GET request and caching-out on subsequent GET requests, passing thru transparently PUT,POST,HEAD,DELETE and COPY requests.

The feature introduces 2 new APIs: Auth and Cache.

## New APIs

There are 2 new APIs for this feature:

Auth API - The cache uses this to validate that a user can access the cached data

Cache API - Adds the ability to override securely Range header, that way Nginx can use its own smart cache on top of S3: <https://www.nginx.com/blog/smart-efficient-byte-range-caching-nginx/> Using this API gives the ability to read ahead objects when clients asking a specific range from the object. On subsequent accesses to the cached object, Nginx will satisfy requests for already-cached ranges from the cache. Uncached ranges will be read from RGW (and cached).

## Auth API

This API Validates a specific authenticated access being made to the cache, using RGW's knowledge of the client credentials and stored access policy. Returns success if the encapsulated request would be granted.

## Cache API

This API is meant to allow changing signed Range headers using a privileged user, cache user.

Creating cache user

```
1. $ radosgw-admin user create --uid=<uid for cache user> --display-name="cache user" --caps="amz-cache=read"
```

This user can send to the RGW the Cache API header `X-Amz-Cache`, this header contains the headers from the original request(before changing the Range header). It means that `X-Amz-Cache` built from several headers. The headers that are building the `X-Amz-Cache` header are separated by char with ASCII code 177 and the header name and value are separated by char ASCII code 178. The RGW will check that the cache user is an authorized user and if it is a cache user, if yes it will use the `X-Amz-Cache` to revalidate that the user has permissions, using the headers from the X-Amz-Cache. During this flow, the RGW will override the Range header.

## Using Nginx with RGW

Download the source of Openresty:

```
1. $ wget https://openresty.org/download/openresty-1.15.8.3.tar.gz
```

git clone the AWS auth Nginx module:

```
1. $ git clone https://github.com/kaltura/nginx-aws-auth-module
```

untar the openresty package:

```
1. $ tar xvzf openresty-1.15.8.3.tar.gz
2. $ cd openresty-1.15.8.3
```

Compile openresty, Make sure that you have pcre lib and openssl lib:

```
1. $ sudo yum install pcre-devel openssl-devel gcc curl zlib-devel nginx
$ ./configure --add-module=<the nginx-aws-auth-module dir> --with-http_auth_request_module --with-
2. http_slice_module --conf-path=/etc/nginx/nginx.conf
3. $ gmake -j $(nproc)
4. $ sudo gmake install
5. $ sudo ln -sf /usr/local/openresty/bin/openresty /usr/bin/nginx
```

Put in-place your Nginx configuration files and edit them according to your environment:

All Nginx conf files are under: <https://github.com/ceph/ceph/tree/master/examples/rgw-cache>

nginx.conf should go to /etc/nginx/nginx.conf

nginx-lua-file.lua should go to /etc/nginx/nginx-lua-file.lua

nginx-default.conf should go to /etc/nginx/conf.d/nginx-default.conf

The parameters that are most likely to require adjustment according to the environment are located in the file nginx-default.conf

Modify the example values of *proxy\_cache\_path* and *max\_size* at:

```
1. proxy_cache_path /data/cache levels=2:2:2 keys_zone=mycache:999m max_size=20G inactive=1d use_temp_path=off;
```

And modify the example *server* values to point to the RGWs URIs:

```
1. server rgw1:8000 max_fails=2 fail_timeout=5s;
2. server rgw2:8000 max_fails=2 fail_timeout=5s;
3. server rgw3:8000 max_fails=2 fail_timeout=5s;
```

It is important to substitute the *access key* and *secret key* located in the nginx.conf with those belong to the user with the amz-cache caps

for example, create the cache user as following:

```
radosgw-admin user create --uid=cacheuser --display-name="cache user" --caps="amz-cache=read" --access-key
1. <access> --secret <secret>
```

It is possible to use Nginx slicing which is a better method for streaming purposes.

For using slice you should use nginx-slicing.conf and not nginx-default.conf

Further information about Nginx slicing:

<https://docs.nginx.com/nginx/admin-guide/content-cache/content-caching/#byte-range-caching>

If you do not want to use the prefetch caching, It is possible to replace nginx-default.conf with nginx-noprefetch.conf Using noprefetch means that if the client is sending range request of 0-4095 and then 0-4096 Nginx will cache those requests separately, So it will need to fetch those requests twice.

Run Nginx(openresty):

```
1. $ sudo systemctl restart nginx
```

## Appendix

**A note about performance:** In certain instances like development environment, disabling the authentication by commenting the following line in nginx-default.conf:

```
1. #auth_request /authentication;
```

may (depending on the hardware) increases the performance significantly as it forgoes the auth API calls to radosgw.

# Ceph Object Gateway Swift API

Ceph supports a RESTful API that is compatible with the basic data access model of the [Swift API](#).

## API

- [Authentication](#)
- [Service Ops](#)
- [Container Ops](#)
- [Object Ops](#)
- [Temp URL Ops](#)
- [Tutorial](#)
- [Java](#)
- [Python](#)
- [Ruby](#)

## Features Support

The following table describes the support status for current Swift functional features:

Feature	Status	Remarks
<b>Authentication</b>	Supported	
<b>Get Account Metadata</b>	Supported	
<b>Swift ACLs</b>	Supported	Supports a subset of Swift ACLs
<b>List Containers</b>	Supported	
<b>Delete Container</b>	Supported	
<b>Create Container</b>	Supported	
<b>Get Container Metadata</b>	Supported	
<b>Update Container Metadata</b>	Supported	
<b>Delete Container Metadata</b>	Supported	
<b>List Objects</b>	Supported	

<b>Static Website</b>	Supported	
<b>Create Object</b>	Supported	
<b>Create Large Object</b>	Supported	
<b>Delete Object</b>	Supported	
<b>Get Object</b>	Supported	
<b>Copy Object</b>	Supported	
<b>Get Object Metadata</b>	Supported	
<b>Update Object Metadata</b>	Supported	
<b>Expiring Objects</b>	Supported	
<b>Temporary URLs</b>	Partial Support	No support for container-level keys
<b>Object Versioning</b>	Partial Support	No support for <code>X-History-Location</code>
<b>CORS</b>	Not Supported	

# Authentication

Swift API requests that require authentication must contain an `X-Storage-Token` authentication token in the request header. The token may be retrieved from RADOS Gateway, or from another authenticator. To obtain a token from RADOS Gateway, you must create a user. For example:

```
1. sudo radosgw-admin user create --subuser="{username}:{subusername}" --uid="{username}"
2. --display-name="{Display Name}" --key-type=swift --secret="{password}" --access=full
```

For details on RADOS Gateway administration, see [radosgw-admin](#).

## Note

For those used to the Swift API this is implementing the Swift auth v1.0 API, as such {username} above is generally equivalent to a Swift account and {subusername} is a user under that account.

## Auth Get

To authenticate a user, make a request containing an `X-Auth-User` and a `X-Auth-Key` in the header.

## Syntax

```
1. GET /auth HTTP/1.1
2. Host: swift.radosgwhost.com
3. X-Auth-User: johndoe
4. X-Auth-Key: R7UU0LFDI2ZI9PRCQ53K
```

## Request Headers

### `X-Auth-User`

#### Description

The key RADOS GW username to authenticate.

#### Type

String

#### Required

Yes

`X-Auth-Key`**Description**

The key associated to a RADOS GW username.

**Type****String****Required****Yes**

## Response Headers

The response from the server should include an `X-Auth-Token` value. The response may also contain a `X-Storage-Url` that provides the `{api version}/{account}` prefix that is specified in other requests throughout the API documentation.

`X-Storage-Token`**Description**

The authorization token for the `X-Auth-User` specified in the request.

**Type****String**`X-Storage-Url`**Description**

The URL and `{api version}/{account}` path for the user.

**Type****String**

A typical response looks like this:

```

1. HTTP/1.1 204 No Content
2. Date: Mon, 16 Jul 2012 11:05:33 GMT
3. Server: swift
4. X-Storage-Url: https://swift.radosgwhost.com/v1/ACCT-12345
5. X-Auth-Token: U01CCC8TahFKlWuv9DB09TWHF0nDjpPElha0kAa
6. Content-Length: 0
7. Content-Type: text/plain; charset=UTF-8

```

# Service Operations

To retrieve data about our Swift-compatible service, you may execute `GET` requests using the `X-Storage-Url` value retrieved during authentication.

## List Containers

A `GET` request that specifies the API version and the account will return a list of containers for a particular user account. Since the request returns a particular user's containers, the request requires an authentication token. The request cannot be made anonymously.

### Syntax

- ```
1. GET /{api version}/{account} HTTP/1.1
2. Host: {fqdn}
3. X-Auth-Token: {auth-token}
```

## Request Parameters

`limit`

Description

Limits the number of results to the specified value.

Type

Integer

Required

No

`format`

Description

Defines the format of the result.

Type

String

Valid Values

`json` | `xml`

Required

No

marker

Description

Returns a list of results greater than the marker value.

Type

String

Required

No

## Response Entities

The response contains a list of containers, or returns with an HTTP 204 response code

account

Description

A list for account information.

Type

Container

container

Description

The list of containers.

Type

Container

name

Description

The name of a container.

Type

String

bytes

Description

The size of the container.

Type

Integer

# Container Operations

A container is a mechanism for storing data objects. An account may have many containers, but container names must be unique. This API enables a client to create a container, set access controls and metadata, retrieve a container's contents, and delete a container. Since this API makes requests related to information in a particular user's account, all requests in this API must be authenticated unless a container's access control is deliberately made publicly accessible (i.e., allows anonymous requests).

## Note

The Amazon S3 API uses the term 'bucket' to describe a data container. When you hear someone refer to a 'bucket' within the Swift API, the term 'bucket' may be construed as the equivalent of the term 'container.'

One facet of object storage is that it does not support hierarchical paths or directories. Instead, it supports one level consisting of one or more containers, where each container may have objects. The RADOS Gateway's Swift-compatible API supports the notion of 'pseudo-hierarchical containers,' which is a means of using object naming to emulate a container (or directory) hierarchy without actually implementing one in the storage system. You may name objects with pseudo-hierarchical names (e.g., photos/buildings/empire-state.jpg), but container names cannot contain a forward slash ( / ) character.

## Create a Container

To create a new container, make a `PUT` request with the API version, account, and the name of the new container. The container name must be unique, must not contain a forward-slash (/) character, and should be less than 256 bytes. You may include access control headers and metadata headers in the request. The operation is idempotent; that is, if you make a request to create a container that already exists, it will return with a HTTP 202 return code, but will not create another container.

## Syntax

1. `PUT /{api version}/{account}/{container} HTTP/1.1`
2. `Host: {fqdn}`
3. `X-Auth-Token: {auth-token}`
4. `X-Container-Read: {comma-separated-uids}`
5. `X-Container-Write: {comma-separated-uids}`
6. `X-Container-Meta-{key}: {value}`

## Headers

**X-Container-Read****Description**

The user IDs with read permissions for the container.

**Type**

Comma-separated string values of user IDs.

**Required**

No

**X-Container-Write****Description**

The user IDs with write permissions for the container.

**Type**

Comma-separated string values of user IDs.

**Required**

No

**X-Container-Meta-{key}****Description**

A user-defined meta data key that takes an arbitrary string value.

**Type**

String

**Required**

No

## HTTP Response

If a container with the same name already exists, and the user is the container owner then the operation will succeed. Otherwise the operation will fail.

**409****Description**

The container already exists under a different user's ownership.

**Status Code**

BucketAlreadyExists

## List a Container's Objects

To list the objects within a container, make a `GET` request with the API version, account, and the name of the container. You can specify query parameters to filter the full list, or leave out the parameters to return a list of the first 10,000 object names stored in the container.

## Syntax

```
1. GET /{api version}/{container} HTTP/1.1
2. Host: {fqdn}
3. X-Auth-Token: {auth-token}
```

## Parameters

format

Description

Defines the format of the result.

Type

String

Valid Values

json | xml

Required

No

prefix

Description

Limits the result set to objects beginning with the specified prefix.

Type

String

Required

No

marker

## Description

Returns a list of results greater than the marker value.

### Type

String

### Required

No

`limit`

## Description

Limits the number of results to the specified value.

### Type

Integer

### Valid Range

0 - 10,000

### Required

No

`delimiter`

## Description

The delimiter between the prefix and the rest of the object name.

### Type

String

### Required

No

`path`

## Description

The pseudo-hierarchical path of the objects.

### Type

String

### Required

NO

`allow_unordered`

Description

Allows the results to be returned unordered to reduce computation overhead. Cannot be used with `delimiter`.

Type

Boolean

Required

NO

Non-Standard Extension

Yes

## Response Entities

`container`

Description

The container.

Type

Container

`object`

Description

An object within the container.

Type

Container

`name`

Description

The name of an object within the container.

Type

String

`hash`

## Description

A hash code of the object's contents.

### Type

String

`last_modified`

## Description

The last time the object's contents were modified.

### Type

Date

`content_type`

## Description

The type of content within the object.

### Type

String

# Update a Container's ACLs

When a user creates a container, the user has read and write access to the container by default. To allow other users to read a container's contents or write to a container, you must specifically enable the user. You may also specify `*` in the `X-Container-Read` or `X-Container-Write` settings, which effectively enables all users to either read from or write to the container. Setting `*` makes the container public. That is it enables anonymous users to either read from or write to the container.

### Note

If you are planning to expose public read ACL functionality for the Swift API, it is strongly recommended to include the Swift account name in the endpoint definition, so as to most closely emulate the behavior of native OpenStack Swift. To do so, set the `ceph.conf` configuration option `rgw swift account in url = true`, and update your Keystone endpoint to the URL suffix `/v1/AUTH_%(tenant_id)s` (instead of just `/v1`).

# Syntax

1. `POST /{api version}/{account}/{container}` `HTTP/1.1`
2. `Host: {fqdn}`
3. `X-Auth-Token: {auth-token}`
4. `X-Container-Read: *`

```
5. X-Container-Write: {uid1}, {uid2}, {uid3}
```

## Request Headers

X-Container-Read

### Description

The user IDs with read permissions for the container.

### Type

Comma-separated string values of user IDs.

### Required

No

X-Container-Write

### Description

The user IDs with write permissions for the container.

### Type

Comma-separated string values of user IDs.

### Required

No

## Add/Update Container Metadata

To add metadata to a container, make a `POST` request with the API version, account, and container name. You must have write permissions on the container to add or update metadata.

## Syntax

```
1. POST /{api version}/{account}/{container} HTTP/1.1
2. Host: {fqdn}
3. X-Auth-Token: {auth-token}
4. X-Container-Meta-Color: red
5. X-Container-Meta-Taste: salty
```

## Request Headers

X-Container-Meta-{key}

## Description

A user-defined meta data key that takes an arbitrary string value.

## Type

String

## Required

No

# Enable Object Versioning for a Container

To enable object versioning a container, make a `POST` request with the API version, account, and container name. You must have write permissions on the container to add or update metadata.

## Note

Object versioning support is not enabled in radosgw by default; you must set `rgw swift versioning enabled = true` in `ceph.conf` to enable this feature.

## Syntax

1. `POST /{api version}/{account}/{container}` `HTTP/1.1`
2. `Host: {fqdn}`
3. `X-Auth-Token: {auth-token}`
4. `X-Versions-Location: {archive-container}`

## Request Headers

### `X-Versions-Location`

## Description

The name of a container (the “archive container”) that will be used to store versions of the objects in the container that the `POST` request is made on (the “current container”). The archive container need not exist at the time it is being referenced, but once `X-Versions-Location` is set on the current container, and object versioning is thus enabled, the archive container must exist before any further objects are updated or deleted in the current container.

## Note

`X-Versions-Location` is the only versioning-related header that radosgw interprets. `X-History-Location`, supported by native OpenStack Swift, is currently not supported by radosgw.

Type

String

Required

No (if this header is passed with an empty value, object versioning on the current container is disabled, but the archive container continues to exist.)

## Delete a Container

To delete a container, make a `DELETE` request with the API version, account, and the name of the container. The container must be empty. If you'd like to check if the container is empty, execute a `HEAD` request against the container. Once you have successfully removed the container, you will be able to reuse the container name.

## Syntax

1. `DELETE /{api version}/{account}/{container} HTTP/1.1`
2. `Host: {fqdn}`
3. `X-Auth-Token: {auth-token}`

## HTTP Response

204

Description

The container was removed.

Status Code

NoContent

# Object Operations

An object is a container for storing data and metadata. A container may have many objects, but the object names must be unique. This API enables a client to create an object, set access controls and metadata, retrieve an object's data and metadata, and delete an object. Since this API makes requests related to information in a particular user's account, all requests in this API must be authenticated unless the container or object's access control is deliberately made publicly accessible (i.e., allows anonymous requests).

## Create/Update an Object

To create a new object, make a `PUT` request with the API version, account, container name and the name of the new object. You must have write permission on the container to create or update an object. The object name must be unique within the container. The `PUT` request is not idempotent, so if you do not use a unique name, the request will update the object. However, you may use pseudo-hierarchical syntax in your object name to distinguish it from another object of the same name if it is under a different pseudo-hierarchical directory. You may include access control headers and metadata headers in the request.

## Syntax

```
1. PUT /{api version}/{account}/{container}/{object} HTTP/1.1
2.   Host: {fqdn}
3.   X-Auth-Token: {auth-token}
```

## Request Headers

### ETag

Description

An MD5 hash of the object's contents. Recommended.

Type

String

Required

No

### Content-Type

Description

The type of content the object contains.

Type

String

Required

No

`Transfer-Encoding`

Description

Indicates whether the object is part of a larger aggregate object.

Type

String

Valid Values

`chunked`

Required

No

## Copy an Object

Copying an object allows you to make a server-side copy of an object, so that you don't have to download it and upload it under another container/name. To copy the contents of one object to another object, you may make either a `PUT` request or a `COPY` request with the API version, account, and the container name. For a `PUT` request, use the destination container and object name in the request, and the source container and object in the request header. For a `Copy` request, use the source container and object in the request, and the destination container and object in the request header. You must have write permission on the container to copy an object. The destination object name must be unique within the container. The request is not idempotent, so if you do not use a unique name, the request will update the destination object. However, you may use pseudo-hierarchical syntax in your object name to distinguish the destination object from the source object of the same name if it is under a different pseudo-hierarchical directory. You may include access control headers and metadata headers in the request.

## Syntax

1. `PUT /{api version}/{account}/{dest-container}/{dest-object}` `HTTP/1.1`
2. `X-Copy-From: {source-container}/{source-object}`
3. `Host: {fqdn}`

```
4. X-Auth-Token: {auth-token}
```

or alternatively:

```
1. COPY /{api version}/{account}/{source-container}/{source-object} HTTP/1.1  
2. Destination: {dest-container}/{dest-object}
```

## Request Headers

X-Copy-From

Description

Used with a `PUT` request to define the source container/object path.

Type

String

Required

Yes, if using `PUT`

Destination

Description

Used with a `COPY` request to define the destination container/object path.

Type

String

Required

Yes, if using `COPY`

If-Modified-Since

Description

Only copies if modified since the date/time of the source object's `last_modified` attribute.

Type

Date

Required

No

If-Unmodified-Since

## Description

Only copies if not modified since the date/time of the source object's `last_modified` attribute.

### Type

Date

Required

No

`Copy-If-Match`

## Description

Copies only if the ETag in the request matches the source object's ETag.

### Type

ETag.

Required

No

`Copy-If-None-Match`

## Description

Copies only if the ETag in the request does not match the source object's ETag.

### Type

ETag.

Required

No

# Delete an Object

To delete an object, make a `DELETE` request with the API version, account, container and object name. You must have write permissions on the container to delete an object within it. Once you have successfully deleted the object, you will be able to reuse the object name.

## Syntax

```
1. DELETE /{api version}/{account}/{container}/{object} HTTP/1.1
2. Host: {fqdn}
```

```
3. X-Auth-Token: {auth-token}
```

## Get an Object

To retrieve an object, make a `GET` request with the API version, account, container and object name. You must have read permissions on the container to retrieve an object within it.

## Syntax

```
1. GET /{api version}/{account}/{container}/{object} HTTP/1.1  
2. Host: {fqdn}  
3. X-Auth-Token: {auth-token}
```

## Request Headers

range

Description

To retrieve a subset of an object's contents, you may specify a byte range.

Type

Date

Required

No

If-Modified-Since

Description

Only copies if modified since the date/time of the source object's `last_modified` attribute.

Type

Date

Required

No

If-Unmodified-Since

Description

Only copies if not modified since the date/time of the source object's `last_modified`

attribute.

Type

Date

Required

No

`Copy-If-Match`

Description

Copies only if the ETag in the request matches the source object's ETag.

Type

ETag.

Required

No

`Copy-If-None-Match`

Description

Copies only if the ETag in the request does not match the source object's ETag.

Type

ETag.

Required

No

## Response Headers

`Content-Range`

Description

The range of the subset of object contents. Returned only if the range header field was specified in the request

## Get Object Metadata

To retrieve an object's metadata, make a `HEAD` request with the API version, account, container and object name. You must have read permissions on the container to retrieve metadata from an object within the container. This request returns the same header

information as the request for the object itself, but it does not return the object's data.

## Syntax

1. HEAD /{api version}/{account}/{container}/{object} HTTP/1.1
2. Host: {fqdn}
3. X-Auth-Token: {auth-token}

## Add/Update Object Metadata

To add metadata to an object, make a `POST` request with the API version, account, container and object name. You must have write permissions on the parent container to add or update metadata.

## Syntax

1. POST /{api version}/{account}/{container}/{object} HTTP/1.1
2. Host: {fqdn}
3. X-Auth-Token: {auth-token}

## Request Headers

`X-Object-Meta-{key}`

Description

A user-defined meta data key that takes an arbitrary string value.

Type

String

Required

No

# Temp URL Operations

To allow temporary access (for eg for GET requests) to objects without the need to share credentials, temp url functionality is supported by swift endpoint of radosgw. For this functionality, initially the value of X-Account-Meta-Temp-URL-Key and optionally X-Account-Meta-Temp-URL-Key-2 should be set. The Temp URL functionality relies on a HMAC-SHA1 signature against these secret keys.

## Note

If you are planning to expose Temp URL functionality for the Swift API, it is strongly recommended to include the Swift account name in the endpoint definition, so as to most closely emulate the behavior of native OpenStack Swift. To do so, set the `ceph.conf` configuration option `rgw swift account in url = true`, and update your Keystone endpoint to the URL suffix `/v1/AUTH_{tenant_id}s` (instead of just `/v1`).

## POST Temp-URL Keys

A `POST` request to the Swift account with the required key will set the secret temp URL key for the account, against which temporary URL access can be provided to accounts. Up to two keys are supported, and signatures are checked against both the keys, if present, so that keys can be rotated without invalidating the temporary URLs.

## Note

Native OpenStack Swift also supports the option to set temporary URL keys at the container level, issuing a `POST` or `PUT` request against a container that sets `X-Container-Meta-Temp-URL-Key` or `X-Container-Meta-Temp-URL-Key-2`. This functionality is not supported in radosgw; temporary URL keys can only be set and used at the account level.

## Syntax

1. `POST /{api version}/{account} HTTP/1.1`
2. `Host: {fqdn}`
3. `X-Auth-Token: {auth-token}`

## Request Headers

### `X-Account-Meta-Temp-URL-Key`

#### Description

A user-defined key that takes an arbitrary string value.

Type

String

Required

Yes

X-Account-Meta-Temp-URL-Key-2

Description

A user-defined key that takes an arbitrary string value.

Type

String

Required

No

## GET Temp-URL Objects

Temporary URL uses a cryptographic HMAC-SHA1 signature, which includes the following elements:

1. The value of the Request method, "GET" for instance
2. The expiry time, in format of seconds since the epoch, ie Unix time
3. The request path starting from "v1" onwards

The above items are normalized with newlines appended between them, and a HMAC is generated using the SHA-1 hashing algorithm against one of the Temp URL Keys posted earlier.

A sample python script to demonstrate the above is given below:

```

1. import hmac
2. from hashlib import sha1
3. from time import time
4.
5. method = 'GET'
6. host = 'https://objectstore.example.com/swift'
7. duration_in_seconds = 300 # Duration for which the url is valid
8. expires = int(time() + duration_in_seconds)
9. path = '/v1/your-bucket/your-object'
10. key = 'secret'
11. hmac_body = '%s\n%s\n%s' % (method, expires, path)
12. sig = hmac.new(key, hmac_body, sha1).hexdigest()
13. rest_uri = "{host}{path}?temp_url_sig={sig}&temp_url_expires={expires}".format(

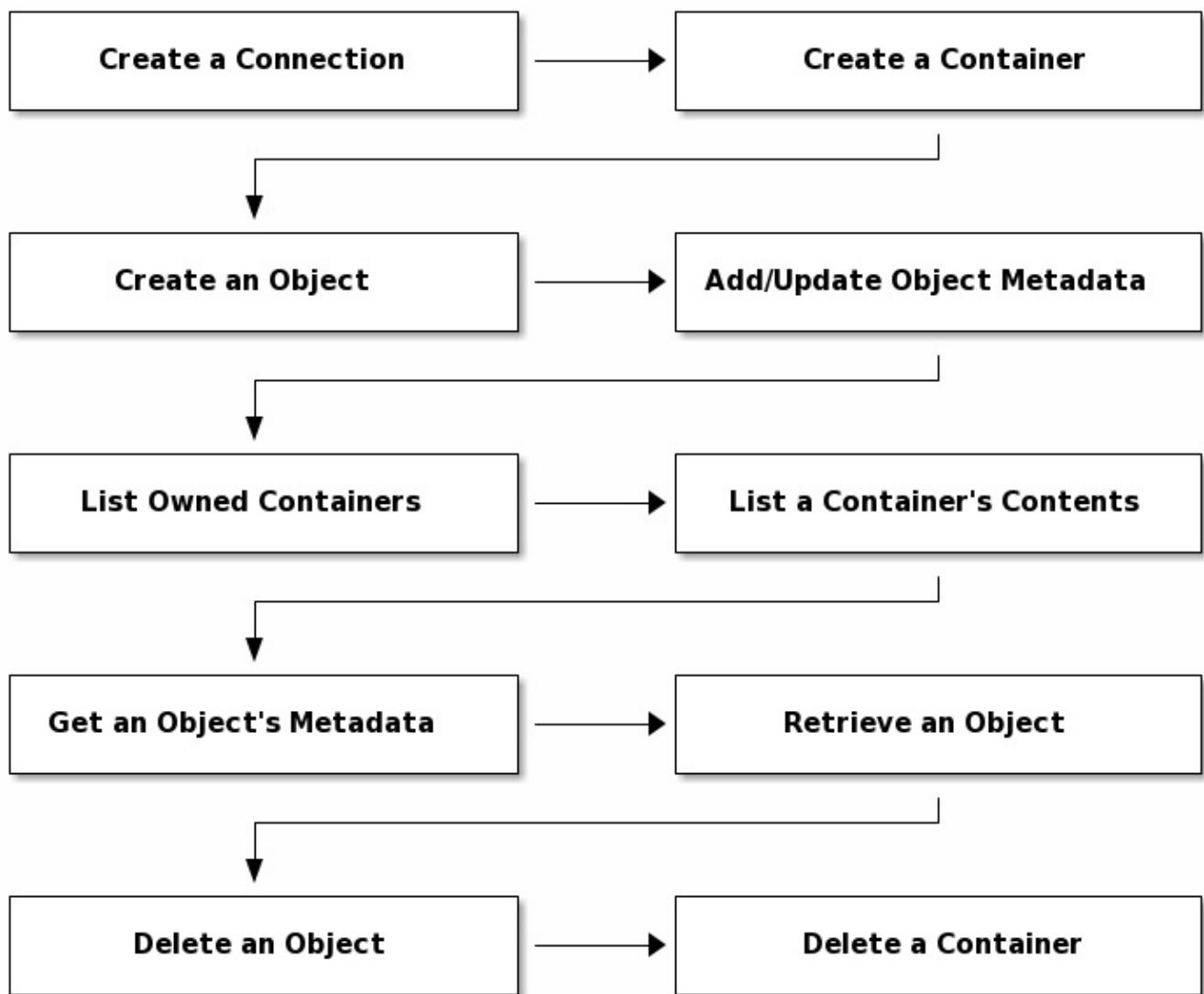
```

```
14.           host=host, path=path, sig=sig, expires=expires)
15. print rest_uri
16.
17. # Example Output
# https://objectstore.example.com/swift/v1/your-bucket/your-object?
18. temp_url_sig=ff4657876227fc6025f04fcf1e82818266d022c6&temp_url_expires=1423200992
```

# Tutorial

The Swift-compatible API tutorials follow a simple container-based object lifecycle. The first step requires you to setup a connection between your client and the RADOS Gateway server. Then, you may follow a natural container and object lifecycle, including adding and retrieving object metadata. See example code for the following languages:

- [Java](#)
- [Python](#)
- [Ruby](#)



# Java Swift Examples

## Setup

The following examples may require some or all of the following Java classes to be imported:

```
1. import org.javaswift.joss.client.factory.AccountConfig;
2. import org.javaswift.joss.client.factory.AccountFactory;
3. import org.javaswift.joss.client.factory.AuthenticationMethod;
4. import org.javaswift.joss.model.Account;
5. import org.javaswift.joss.model.Container;
6. import org.javaswift.joss.model.StoredObject;
7. import java.io.File;
8. import java.io.IOException;
9. import java.util.*;
```

## Create a Connection

This creates a connection so that you can interact with the server:

```
1. String username = "USERNAME";
2. String password = "PASSWORD";
3. String authUrl = "https://radosgw.endpoint/auth/1.0";
4.
5. AccountConfig config = new AccountConfig();
6. config.setUsername(username);
7. config.setPassword(password);
8. config.setAuthUrl(authUrl);
9. config.setAuthenticationMethod(AuthenticationMethod.BASIC);
10. Account account = new AccountFactory(config).createAccount();
```

## Create a Container

This creates a new container called `my-new-container` :

```
1. Container container = account.getContainer("my-new-container");
2. container.create();
```

## Create an Object

This creates an object `foo.txt` from the file named `foo.txt` in the container `my-new-container` :

```

1. Container container = account.getContainer("my-new-container");
2. StoredObject object = container.getObject("foo.txt");
3. object.uploadObject(new File("foo.txt"));

```

## Add/Update Object Metadata

This adds the metadata key-value pair `key : value` to the object named `foo.txt` in the container `my-new-container` :

```

1. Container container = account.getContainer("my-new-container");
2. StoredObject object = container.getObject("foo.txt");
3. Map<String, Object> metadata = new TreeMap<String, Object>();
4. metadata.put("key", "value");
5. object.setMetadata(metadata);

```

## List Owned Containers

This gets a list of Containers that you own. This also prints out the container name.

```

1. Collection<Container> containers = account.list();
2. for (Container currentContainer : containers) {
3.     System.out.println(currentContainer.getName());
4. }

```

The output will look something like this:

```

1. mahbuckat1
2. mahbuckat2
3. mahbuckat3

```

## List a Container's Content

This gets a list of objects in the container `my-new-container` ; and, it also prints out each object's name, the file size, and last modified date:

```

1. Container container = account.getContainer("my-new-container");
2. Collection<StoredObject> objects = container.list();
3. for (StoredObject currentObject : objects) {
4.     System.out.println(currentObject.getName());
5. }

```

The output will look something like this:

```

1. myphoto1.jpg

```

```
2. myphoto2.jpg
```

## Retrieve an Object's Metadata

This retrieves metadata and gets the MIME type for an object named `foo.txt` in a container named `my-new-container` :

```
1. Container container = account.getContainer("my-new-container");
2. StoredObject object = container.getObject("foo.txt");
3. Map<String, Object> returnedMetadata = object.getMetadata();
4. for (String name : returnedMetadata.keySet()) {
5.     System.out.println("META / "+name+": "+returnedMetadata.get(name));
6. }
```

## Retrieve an Object

This downloads the object `foo.txt` in the container `my-new-container` and saves it in `./outfile.txt` :

```
1. Container container = account.getContainer("my-new-container");
2. StoredObject object = container.getObject("foo.txt");
3. object.downloadObject(new File("outfile.txt"));
```

## Delete an Object

This deletes the object `goodbye.txt` in the container "my-new-container":

```
1. Container container = account.getContainer("my-new-container");
2. StoredObject object = container.getObject("foo.txt");
3. object.delete();
```

## Delete a Container

This deletes a container named "my-new-container":

```
1. Container container = account.getContainer("my-new-container");
2. container.delete();
```

### Note

The container must be empty! Otherwise it won't work!

# Python Swift Examples

## Create a Connection

This creates a connection so that you can interact with the server:

```
1. import swiftclient
2. user = 'account_name:username'
3. key = 'your_api_key'
4.
5. conn = swiftclient.Connection(
6.     user=user,
7.     key=key,
8.     authurl='https://objects.dreamhost.com/auth',
9. )
```

## Create a Container

This creates a new container called `my-new-container` :

```
1. container_name = 'my-new-container'
2. conn.put_container(container_name)
```

## Create an Object

This creates a file `hello.txt` from the file named `my_hello.txt` :

```
1. with open('hello.txt', 'r') as hello_file:
2.     conn.put_object(container_name, 'hello.txt',
3.                      contents= hello_file.read(),
4.                      content_type='text/plain')
```

## List Owned Containers

This gets a list of containers that you own, and prints out the container name:

```
1. for container in conn.get_account()[1]:
2.     print container['name']
```

The output will look something like this:

```
1. mahbuckat1
```

```
2. mahbuckat2
3. mahbuckat3
```

## List a Container's Content

This gets a list of objects in the container, and prints out each object's name, the file size, and last modified date:

```
1. for data in conn.get_container(container_name)[1]:
2.     print '{0}\t{1}\t{2}'.format(data['name'], data['bytes'], data['last_modified'])
```

The output will look something like this:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Retrieve an Object

This downloads the object `hello.txt` and saves it in `./my_hello.txt` :

```
1. obj_tuple = conn.get_object(container_name, 'hello.txt')
2. with open('my_hello.txt', 'w') as my_hello:
3.     my_hello.write(obj_tuple[1])
```

## Delete an Object

This deletes the object `hello.txt` :

```
1. conn.delete_object(container_name, 'hello.txt')
```

## Delete a Container

### Note

The container must be empty! Otherwise the request won't work!

```
1. conn.delete_container(container_name)
```

# Ruby Swift Examples

## Create a Connection

This creates a connection so that you can interact with the server:

```
1. require 'cloudfiles'
2. username = 'account_name:user_name'
3. api_key  = 'your_secret_key'
4.
5. conn = CloudFiles::Connection.new(
6.     :username => username,
7.     :api_key   => api_key,
8.     :auth_url => 'http://objects.dreamhost.com/auth'
9. )
```

## Create a Container

This creates a new container called `my-new-container`

```
1. container = conn.create_container('my-new-container')
```

## Create an Object

This creates a file `hello.txt` from the file named `my_hello.txt`

```
1. obj = container.create_object('hello.txt')
2. obj.load_from_filename('./my_hello.txt')
3. obj.content_type = 'text/plain'
```

## List Owned Containers

This gets a list of Containers that you own, and also prints out the container name:

```
1. conn.containers.each do |container|
2.     puts container
3. end
```

The output will look something like this:

```
1. mahbuckat1
2. mahbuckat2
```

```
3. mahbuckat3
```

## List a Container's Contents

This gets a list of objects in the container, and prints out each object's name, the file size, and last modified date:

```
1. require 'date' # not necessary in the next version
2.
3. container.objects_detail.each do |name, data|
4.   puts "#{name}\t#{data[:bytes]}\t#{data[:last_modified]}"
5. end
```

The output will look something like this:

```
1. myphoto1.jpg 251262 2011-08-08T21:35:48.000Z
2. myphoto2.jpg 262518 2011-08-08T21:38:01.000Z
```

## Retrieve an Object

This downloads the object `hello.txt` and saves it in `./my_hello.txt` :

```
1. obj = container.object('hello.txt')
2. obj.save_to_filename('./my_hello.txt')
```

## Delete an Object

This deletes the object `goodbye.txt` :

```
1. container.delete_object('goodbye.txt')
```

## Delete a Container

### Note

The container must be empty! Otherwise the request won't work!

```
1. container.delete_container('my-new-container')
```

# Admin Operations

An admin API request will be done on a URI that starts with the configurable 'admin' resource entry point. Authorization for the admin API duplicates the S3 authorization mechanism. Some operations require that the user holds special administrative capabilities. The response entity type (XML or JSON) may be specified as the 'format' option in the request and defaults to JSON if not specified.

## Get Usage

Request bandwidth usage information.

Note: this feature is disabled by default, can be enabled by setting `rgw enable usage log = true` in the appropriate section of `ceph.conf`. For changes in `ceph.conf` to take effect, `radosgw` process restart is needed.

`caps`

`usage=read`

## Syntax

1. `GET /{admin}/usage?format=json HTTP/1.1`
2. `Host: {fqdn}`

## Request Parameters

`uid`

Description

The user for which the information is requested. If not specified will apply to all users.

Type

String

Example

`foo_user`

Required

No

`start`

## Description

Date and (optional) time that specifies the start time of the requested data.

## Type

String

## Example

```
2012-09-25 16:00:00
```

## Required

No

```
end
```

## Description

Date and (optional) time that specifies the end time of the requested data (non-inclusive).

## Type

String

## Example

```
2012-09-25 16:00:00
```

## Required

No

```
show-entries
```

## Description

Specifies whether data entries should be returned.

## Type

Boolean

## Example

```
True [True]
```

## Required

No

```
show-summary
```

## Description

Specifies whether data summary should be returned.

Type

Boolean

Example

True [True]

Required

No

## Response Entities

If successful, the response contains the requested information.

usage

Description

A container for the usage information.

Type

Container

entries

Description

A container for the usage entries information.

Type

Container

user

Description

A container for the user data information.

Type

Container

owner

Description

The name of the user that owns the buckets.

Type

Type

String

Description

The bucket name.

Type

String

time

Description

Time lower bound for which data is being specified (rounded to the beginning of the first relevant hour).

Type

String

epoch

Description

The time specified in seconds since 1/1/1970.

Type

String

categories

Description

A container for stats categories.

Type

Container

entry

Description

A container for stats entry.

Type

Container

category

Description

Name of request category for which the stats are provided.

Type

String

`bytes_sent`

Description

Number of bytes sent by the RADOS Gateway.

Type

Integer

`bytes_received`

Description

Number of bytes received by the RADOS Gateway.

Type

Integer

`ops`

Description

Number of operations.

Type

Integer

`successful_ops`

Description

Number of successful operations.

Type

Integer

`summary`

Description

A container for stats summary.

Type

Container

`total`

#### Description

A container for stats summary aggregated total.

#### Type

Container

## Special Error Responses

TBD.

## Trim Usage

Remove usage information. With no dates specified, removes all usage information.

Note: this feature is disabled by default, can be enabled by setting `rgw enable usage log = true` in the appropriate section of ceph.conf. For changes in ceph.conf to take effect, radosgw process restart is needed.

caps

`usage=write`

## Syntax

1. `DELETE /{admin}/usage?format=json HTTP/1.1`
2. `Host: {fqdn}`

## Request Parameters

`uid`

#### Description

The user for which the information is requested. If not specified will apply to all users.

#### Type

String

#### Example

`foo_user`

#### Required

No

start

#### Description

Date and (optional) time that specifies the start time of the requested data.

Type

String

#### Example

2012-09-25 16:00:00

Required

No

end

#### Description

Date and (optional) time that specifies the end time of the requested data (none inclusive).

Type

String

#### Example

2012-09-25 16:00:00

Required

No

remove-all

#### Description

Required when uid is not specified, in order to acknowledge multi user data removal.

Type

Boolean

#### Example

True [False]

Required

No

# Special Error Responses

TBD.

## Get User Info

Get user information.

caps

users=read

### Syntax

1. GET /{admin}/user?format=json HTTP/1.1
2. Host: {fqdn}

## Request Parameters

uid

Description

The user for which the information is requested.

Type

String

Example

foo\_user

Required

Yes

## Response Entities

If successful, the response contains the user information.

user

Description

A container for the user data information.

Type

Container

`user_id`

#### Description

The user id.

#### Type

String

#### Parent

`user`

`display_name`

#### Description

Display name for the user.

#### Type

String

#### Parent

`user`

`suspended`

#### Description

True if the user is suspended.

#### Type

Boolean

#### Parent

`user`

`max_buckets`

#### Description

The maximum number of buckets to be owned by the user.

#### Type

Integer

#### Parent

`user`

`subusers`

## Description

Subusers associated with this user account.

## Type

Container

## Parent

`user`

`keys`

## Description

S3 keys associated with this user account.

## Type

Container

## Parent

`user`

`swift_keys`

## Description

Swift keys associated with this user account.

## Type

Container

## Parent

`user`

`caps`

## Description

User capabilities.

## Type

Container

## Parent

`user`

# Special Error Responses

None.

## Create User

Create a new user. By default, a S3 key pair will be created automatically and returned in the response. If only one of `access-key` or `secret-key` is provided, the omitted key will be automatically generated. By default, a generated key is added to the keyring without replacing an existing key pair. If `access-key` is specified and refers to an existing key owned by the user then it will be modified.

New in version Luminous.

A `tenant` may either be specified as a part of uid or as an additional request param.

`caps`

`users=write`

## Syntax

```
1. PUT /{admin}/user?format=json HTTP/1.1
2. Host: {fqdn}
```

## Request Parameters

`uid`

Description

The user ID to be created.

Type

String

Example

`foo_user`

Required

Yes

A tenant name may also be specified as a part of `uid`, by following the syntax `tenant$user`, refer to [Multitenancy](#) for more details.

`display-name`

Description

The display name of the user to be created.

Type

String

Example

```
foo user
```

Required

Yes

```
email
```

Description

The email address associated with the user.

Type

String

Example

```
foo@bar.com
```

Required

No

```
key-type
```

Description

Key type to be generated, options are: swift, s3 (default).

Type

String

Example

```
s3 [ s3 ]
```

Required

No

```
access-key
```

Description

Specify access key.

Type

String

Example

```
ABCD0EF12GHIJ2K34LMN
```

Required

No

```
secret-key
```

Description

Specify secret key.

Type

String

Example

```
0AbCDEFg1h2i34Jk1M5nop6QrSTUV+WxyzaBC7D8
```

Required

No

```
user-caps
```

Description

User capabilities.

Type

String

Example

```
usage=read, write; users=read
```

Required

No

```
generate-key
```

Description

Generate a new key pair and add to the existing keyring.

Type

Boolean

Example

True [True]

Required

No

max-buckets

Description

Specify the maximum number of buckets the user can own.

Type

Integer

Example

500 [1000]

Required

No

suspended

Description

Specify whether the user should be suspended.

Type

Boolean

Example

False [False]

Required

No

New in version Jewel.

tenant

Description

the Tenant under which a user is a part of.

Type

string

Example

tenant1

Required

No

## Response Entities

If successful, the response contains the user information.

user

Description

A container for the user data information.

Type

Container

tenant

Description

The tenant which user is a part of.

Type

String

Parent

user

user\_id

Description

The user id.

Type

String

Parent

user

display\_name

Description

Display name for the user.

Type

String

Parent

  user

  suspended

Description

True if the user is suspended.

Type

Boolean

Parent

  user

  max\_buckets

Description

The maximum number of buckets to be owned by the user.

Type

Integer

Parent

  user

  subusers

Description

Subusers associated with this user account.

Type

Container

Parent

  user

  keys

Description

S3 keys associated with this user account.

Type

Container

Parent

user

swift\_keys

Description

Swift keys associated with this user account.

Type

Container

Parent

user

caps

Description

User capabilities.

Type

Container

Parent

user

## Special Error Responses

UserExists

Description

Attempt to create existing user.

Code

409 Conflict

InvalidAccessKey

Description

Invalid access key specified.

Code

## 400 Bad Request

`InvalidKeyType`

### Description

Invalid key type specified.

### Code

## 400 Bad Request

`InvalidSecretKey`

### Description

Invalid secret key specified.

### Code

## 400 Bad Request

`InvalidKeyType`

### Description

Invalid key type specified.

### Code

## 400 Bad Request

`KeyExists`

### Description

Provided access key exists and belongs to another user.

### Code

## 409 Conflict

`EmailExists`

### Description

Provided email address exists.

### Code

## 409 Conflict

`InvalidCapability`

### Description

Attempt to grant invalid admin capability.

Code

400 Bad Request

## Modify User

Modify a user.

caps

users=write

## Syntax

1. POST /{admin}/user?format=json HTTP/1.1
2. Host: {fqdn}

## Request Parameters

uid

Description

The user ID to be modified.

Type

String

Example

foo\_user

Required

Yes

display-name

Description

The display name of the user to be modified.

Type

String

Example

foo user

Required

No

email

Description

The email address to be associated with the user.

Type

String

Example

foo@bar.com

Required

No

generate-key

Description

Generate a new key pair and add to the existing keyring.

Type

Boolean

Example

True [False]

Required

No

access-key

Description

Specify access key.

Type

String

Example

ABCD0EF12GHIJ2K34LMN

## Required

No

secret-key

### Description

Specify secret key.

Type

String

### Example

0AbCDEFg1h2i34Jk1M5nop6QrSTUV+WxyzabC7D8

## Required

No

key-type

### Description

Key type to be generated, options are: swift, s3 (default).

Type

String

### Example

s3

## Required

No

user-caps

### Description

User capabilities.

Type

String

### Example

usage=read, write; users=read

## Required

No

max-buckets

#### Description

Specify the maximum number of buckets the user can own.

Type

Integer

#### Example

500 [1000]

Required

No

suspended

#### Description

Specify whether the user should be suspended.

Type

Boolean

#### Example

False [False]

Required

No

op-mask

#### Description

The op-mask of the user to be modified.

Type

String

#### Example

read, write, delete, \*

Required

No

# Response Entities

If successful, the response contains the user information.

`user`

Description

A container for the user data information.

Type

Container

`user_id`

Description

The user id.

Type

String

Parent

`user`

`display_name`

Description

Display name for the user.

Type

String

Parent

`user`

`suspended`

Description

True if the user is suspended.

Type

Boolean

Parent

`user`

max\_buckets

#### Description

The maximum number of buckets to be owned by the user.

#### Type

Integer

#### Parent

user

subusers

#### Description

Subusers associated with this user account.

#### Type

Container

#### Parent

user

keys

#### Description

S3 keys associated with this user account.

#### Type

Container

#### Parent

user

swift\_keys

#### Description

Swift keys associated with this user account.

#### Type

Container

#### Parent

user

caps

## Description

User capabilities.

## Type

Container

## Parent

user

# Special Error Responses

InvalidAccessKey

## Description

Invalid access key specified.

## Code

400 Bad Request

InvalidKeyType

## Description

Invalid key type specified.

## Code

400 Bad Request

InvalidSecretKey

## Description

Invalid secret key specified.

## Code

400 Bad Request

KeyExists

## Description

Provided access key exists and belongs to another user.

## Code

409 Conflict

EmailExists

## Description

Provided email address exists.

## Code

409 Conflict

`InvalidCapability`

## Description

Attempt to grant invalid admin capability.

## Code

400 Bad Request

# Remove User

Remove an existing user.

caps

users=write

## Syntax

```
1. DELETE /{admin}/user?format=json HTTP/1.1
2. Host: {fqdn}
```

# Request Parameters

`uid`

## Description

The user ID to be removed.

## Type

String

## Example

`foo_user`

## Required

Yes.

`purge-data`

#### Description

When specified the buckets and objects belonging to the user will also be removed.

#### Type

Boolean

#### Example

True

Required

No

## Response Entities

None

## Special Error Responses

None.

## Create Subuser

Create a new subuser (primarily useful for clients using the Swift API). Note that in general for a subuser to be useful, it must be granted permissions by specifying

`access`. As with user creation if `subuser` is specified without `secret`, then a secret key will be automatically generated.

`caps`

`users=write`

## Syntax

1. `PUT /{admin}/user?subuser&format=json` HTTP/1.1
2. `Host {fqdn}`

## Request Parameters

`uid`

#### Description

The user ID under which a subuser is to be created.

Type

String

Example

```
foo_user
```

Required

Yes

```
subuser
```

Description

Specify the subuser ID to be created.

Type

String

Example

```
sub_foo
```

Required

Yes

```
secret-key
```

Description

Specify secret key.

Type

String

Example

```
0AbCDEFg1h2i34Jk1M5nop6QrSTUV+WxyzabC7D8
```

Required

No

```
key-type
```

Description

Key type to be generated, options are: swift (default), s3.

Type

String

Example

```
swift [ swift ]
```

Required

No

```
access
```

Description

Set access permissions for sub-user, should be one of `read, write, readwrite, full`.

Type

String

Example

```
read
```

Required

No

```
generate-secret
```

Description

Generate the secret key.

Type

Boolean

Example

```
True [False]
```

Required

No

## Response Entities

If successful, the response contains the subuser information.

```
subusers
```

Description

Subusers associated with the user account.

Type

Container

`id`

Description

Subuser id.

Type

String

Parent

`subusers`

`permissions`

Description

Subuser access to user account.

Type

String

Parent

`subusers`

## Special Error Responses

`SubuserExists`

Description

Specified subuser exists.

Code

409 Conflict

`InvalidKeyType`

Description

Invalid key type specified.

Code

400 Bad Request

`InvalidSecretKey`

## Description

Invalid secret key specified.

## Code

400 Bad Request

`InvalidAccess`

## Description

Invalid subuser access specified.

## Code

400 Bad Request

# Modify Subuser

Modify an existing subuser

caps

users=write

## Syntax

1. POST `/{admin}/user?subuser&format=json` HTTP/1.1
2. Host `{fqdn}`

## Request Parameters

`uid`

### Description

The user ID under which the subuser is to be modified.

### Type

String

### Example

`foo_user`

### Required

Yes

subuser

#### Description

The subuser ID to be modified.

#### Type

String

#### Example

sub\_foo

#### Required

Yes

generate-secret

#### Description

Generate a new secret key for the subuser, replacing the existing key.

#### Type

Boolean

#### Example

True [False]

#### Required

No

secret

#### Description

Specify secret key.

#### Type

String

#### Example

0AbCDEFg1h2i34Jk1M5nop6QrSTUV+WxyzabC7D8

#### Required

No

key-type

### Description

Key type to be generated, options are: swift (default), s3 .

### Type

String

### Example

```
swift [ swift ]
```

### Required

No

```
access
```

### Description

Set access permissions for sub-user, should be one of `read, write, readwrite, full` .

### Type

String

### Example

```
read
```

### Required

No

## Response Entities

If successful, the response contains the subuser information.

```
subusers
```

### Description

Subusers associated with the user account.

### Type

Container

```
id
```

### Description

Subuser id.

### Type

String

Parent

subusers

permissions

Description

Subuser access to user account.

Type

String

Parent

subusers

## Special Error Responses

InvalidKeyType

Description

Invalid key type specified.

Code

400 Bad Request

InvalidSecretKey

Description

Invalid secret key specified.

Code

400 Bad Request

InvalidAccess

Description

Invalid subuser access specified.

Code

400 Bad Request

## Remove Subuser

Remove an existing subuser

caps

users=write

## Syntax

- ```
1. DELETE /{admin}/user?subuser&format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

`uid`

Description

The user ID under which the subuser is to be removed.

Type

String

Example

`foo_user`

Required

Yes

`subuser`

Description

The subuser ID to be removed.

Type

String

Example

`sub_foo`

Required

Yes

`purge-keys`

Description

Remove keys belonging to the subuser.

Type

Boolean

Example

True [True]

Required

No

## Response Entities

None.

## Special Error Responses

None.

## Create Key

Create a new key. If a `subuser` is specified then by default created keys will be swift type. If only one of `access-key` or `secret-key` is provided the committed key will be automatically generated, that is if only `secret-key` is specified then `access-key` will be automatically generated. By default, a generated key is added to the keyring without replacing an existing key pair. If `access-key` is specified and refers to an existing key owned by the user then it will be modified. The response is a container listing all keys of the same type as the key created. Note that when creating a swift key, specifying the option `access-key` will have no effect. Additionally, only one swift key may be held by each user or subuser.

caps

users=write

## Syntax

```
1. PUT /{admin}/user?key&format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

`uid`

## Description

The user ID to receive the new key.

## Type

String

## Example

`foo_user`

## Required

Yes

`subuser`

## Description

The subuser ID to receive the new key.

## Type

String

## Example

`sub_foo`

## Required

No

`key-type`

## Description

Key type to be generated, options are: swift, s3 (default).

## Type

String

## Example

`s3` [ `s3` ]

## Required

No

`access-key`

## Description

Specify the access key.

Type

String

Example

```
AB01C2D3EF45G6H7IJ8K
```

Required

No

```
secret-key
```

Description

Specify the secret key.

Type

String

Example

```
0ab/CdeFGhij1klmnopqRSTUv1WxyZabcDEFgHij
```

Required

No

```
generate-key
```

Description

Generate a new key pair and add to the existing keyring.

Type

Boolean

Example

```
True [ True ]
```

Required

No

## Response Entities

```
keys
```

Description

Keys of type created associated with this user account.

Type

Container

`user`

Description

The user account associated with the key.

Type

String

Parent

`keys`

`access-key`

Description

The access key.

Type

String

Parent

`keys`

`secret-key`

Description

The secret key

Type

String

Parent

`keys`

## Special Error Responses

`InvalidAccessKey`

Description

Invalid access key specified.

Code

400 Bad Request

`InvalidKeyType`

Description

Invalid key type specified.

Code

400 Bad Request

`InvalidSecretKey`

Description

Invalid secret key specified.

Code

400 Bad Request

`InvalidKeyType`

Description

Invalid key type specified.

Code

400 Bad Request

`KeyExists`

Description

Provided access key exists and belongs to another user.

Code

409 Conflict

## Remove Key

Remove an existing key.

caps

`users=write`

## Syntax

```
1. DELETE /{admin}/user?key&format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

access-key

### Description

The S3 access key belonging to the S3 key pair to remove.

### Type

String

### Example

AB01C2D3EF45G6H7IJ8K

### Required

Yes

uid

### Description

The user to remove the key from.

### Type

String

### Example

foo\_user

### Required

No

subuser

### Description

The subuser to remove the key from.

### Type

String

### Example

sub\_foo

## Required

No

`key-type`

## Description

Key type to be removed, options are: swift, s3. NOTE: Required to remove swift key.

## Type

String

## Example

`swift`

## Required

No

# Special Error Responses

None.

# Response Entities

None.

# Get Bucket Info

Get information about a subset of the existing buckets. If `uid` is specified without `bucket` then all buckets belonging to the user will be returned. If `bucket` alone is specified, information for that particular bucket will be retrieved.

`caps`

`buckets=read`

# Syntax

1. GET `/{admin}/bucket?format=json` HTTP/1.1
2. Host `{fqdn}`

# Request Parameters

`bucket`

## Description

The bucket to return info on.

### Type

String

### Example

```
foo_bucket
```

### Required

No

```
uid
```

## Description

The user to retrieve bucket information for.

### Type

String

### Example

```
foo_user
```

### Required

No

```
stats
```

## Description

Return bucket statistics.

### Type

Boolean

### Example

```
True [False]
```

### Required

No

## Response Entities

If successful the request returns a buckets container containing the desired bucket

information.

`stats`

Description

Per bucket information.

Type

Container

`buckets`

Description

Contains a list of one or more bucket containers.

Type

Container

`bucket`

Description

Container for single bucket information.

Type

Container

Parent

`buckets`

`name`

Description

The name of the bucket.

Type

String

Parent

`bucket`

`pool`

Description

The pool the bucket is stored in.

Type

String

Parent

bucket

id

Description

The unique bucket id.

Type

String

Parent

bucket

marker

Description

Internal bucket tag.

Type

String

Parent

bucket

owner

Description

The user id of the bucket owner.

Type

String

Parent

bucket

usage

Description

Storage usage information.

Type

## Container

### Parent

bucket

index

#### Description

Status of bucket index.

#### Type

String

### Parent

bucket

## Special Error Responses

IndexRepairFailed

#### Description

Bucket index repair failed.

#### Code

409 Conflict

## Check Bucket Index

Check the index of an existing bucket. NOTE: to check multipart object accounting with

check-objects , fix must be set to True.

#### caps

buckets=write

## Syntax

```
1. GET /{admin}/bucket?index&format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

bucket

#### Description

The bucket to return info on.

Type

String

Example

```
foo_bucket
```

Required

Yes

```
check-objects
```

Description

Check multipart object accounting.

Type

Boolean

Example

```
True [False]
```

Required

No

```
fix
```

Description

Also fix the bucket index when checking.

Type

Boolean

Example

```
False [False]
```

Required

No

## Response Entities

```
index
```

Description

Status of bucket index.

Type

String

## Special Error Responses

`IndexRepairFailed`

Description

Bucket index repair failed.

Code

409 Conflict

## Remove Bucket

Delete an existing bucket.

caps

buckets=write

## Syntax

```
1. DELETE /{admin}/bucket?format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

`bucket`

Description

The bucket to remove.

Type

String

Example

`foo_bucket`

Required

Yes

`purge-objects`

#### Description

Remove a buckets objects before deletion.

#### Type

Boolean

#### Example

True [False]

#### Required

No

## Response Entities

None.

## Special Error Responses

`BucketNotEmpty`

#### Description

Attempted to delete non-empty bucket.

#### Code

409 Conflict

`ObjectRemovalFailed`

#### Description

Unable to remove objects.

#### Code

409 Conflict

## Unlink Bucket

Unlink a bucket from a specified user. Primarily useful for changing bucket ownership.

caps

buckets=write

## Syntax

```
1. POST /{admin}/bucket?format=json HTTP/1.1  
2. Host {fqdn}
```

## Request Parameters

`bucket`

Description

The bucket to unlink.

Type

String

Example

`foo_bucket`

Required

Yes

`uid`

Description

The user ID to unlink the bucket from.

Type

String

Example

`foo_user`

Required

Yes

## Response Entities

None.

## Special Error Responses

`BucketUnlinkFailed`

Description

Unable to unlink bucket from specified user.

Code

409 Conflict

## Link Bucket

Link a bucket to a specified user, unlinking the bucket from any previous user.

caps

buckets=write

## Syntax

- ```
1. PUT /{admin}/bucket?format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

bucket

Description

The bucket name to unlink.

Type

String

Example

foo\_bucket

Required

Yes

bucket-id

Description

The bucket id to unlink.

Type

String

Example

dev.6607669.420

Required

No

uid

Description

The user ID to link the bucket to.

Type

String

Example

foo\_user

Required

Yes

## Response Entities

bucket

Description

Container for single bucket information.

Type

Container

name

Description

The name of the bucket.

Type

String

Parent

bucket

pool

Description

The pool the bucket is stored in.

Type

String

Parent

bucket

id

Description

The unique bucket id.

Type

String

Parent

bucket

marker

Description

Internal bucket tag.

Type

String

Parent

bucket

owner

Description

The user id of the bucket owner.

Type

String

Parent

bucket

usage

Description

Storage usage information.

Type

## Container

### Parent

bucket

index

#### Description

Status of bucket index.

#### Type

String

### Parent

bucket

## Special Error Responses

BucketUnlinkFailed

#### Description

Unable to unlink bucket from specified user.

#### Code

409 Conflict

BucketLinkFailed

#### Description

Unable to link bucket to specified user.

#### Code

409 Conflict

## Remove Object

Remove an existing object. NOTE: Does not require owner to be non-suspended.

#### caps

buckets=write

## Syntax

```
1. DELETE /{admin}/bucket?object&format=json HTTP/1.1
```

2. Host {fqdn}

## Request Parameters

bucket

Description

The bucket containing the object to be removed.

Type

String

Example

foo\_bucket

Required

Yes

object

Description

The object to remove.

Type

String

Example

foo.txt

Required

Yes

## Response Entities

None.

## Special Error Responses

NoSuchObject

Description

Specified object does not exist.

Code

## 404 Not Found

`ObjectRemovalFailed`

## Description

Unable to remove objects.

## Code

409 Conflict

## Get Bucket or Object Policy

Read the policy of an object or bucket.

caps

buckets=read

### Syntax

```
1. GET /{admin}/bucket?policy&format=json HTTP/1.1
2. Host {fqdn}
```

## Request Parameters

`bucket`

## Description

The bucket to read the policy from.

## Type

String

## Example

`foo_bucket`

## Required

Yes

`object`

## Description

The object to read the policy from.

Type

String

Example

```
foo.txt
```

Required

No

## Response Entities

If successful, returns the object or bucket policy

```
policy
```

Description

Access control policy.

Type

Container

## Special Error Responses

```
IncompleteBody
```

Description

Either bucket was not specified for a bucket policy request or bucket and object were not specified for an object policy request.

Code

400 Bad Request

## Add A User Capability

Add an administrative capability to a specified user.

caps

users=write

## Syntax

1. `PUT /{admin}/user?caps&format=json HTTP/1.1`
2. `Host {fqdn}`

## Request Parameters

`uid`

### Description

The user ID to add an administrative capability to.

### Type

String

### Example

`foo_user`

### Required

Yes

`user-caps`

### Description

The administrative capability to add to the user.

### Type

String

### Example

`usage=read,write;user=write`

### Required

Yes

## Response Entities

If successful, the response contains the user's capabilities.

`user`

### Description

A container for the user data information.

### Type

Container

### Parent

user

user\_id

#### Description

The user id.

#### Type

String

#### Parent

user

caps

#### Description

User capabilities.

#### Type

Container

#### Parent

user

## Special Error Responses

InvalidCapability

#### Description

Attempt to grant invalid admin capability.

#### Code

400 Bad Request

## Example Request

```
1. PUT /{admin}/user?caps&user-caps=usage=read,write;user=write&format=json HTTP/1.1
2. Host: {fqdn}
3. Content-Type: text/plain
4. Authorization: {your-authorization-token}
```

## Remove A User Capability

Remove an administrative capability from a specified user.

caps

users=write

## Syntax

```
1. DELETE /{admin}/user?caps&format=json HTTP/1.1  
2. Host {fqdn}
```

## Request Parameters

uid

Description

The user ID to remove an administrative capability from.

Type

String

Example

foo\_user

Required

Yes

user-caps

Description

The administrative capabilities to remove from the user.

Type

String

Example

usage=read, write

Required

Yes

## Response Entities

If successful, the response contains the user's capabilities.

user

## Description

A container for the user data information.

## Type

Container

## Parent

`user`

`user_id`

## Description

The user id.

## Type

String

## Parent

`user`

`caps`

## Description

User capabilities.

## Type

Container

## Parent

`user`

# Special Error Responses

`InvalidCapability`

## Description

Attempt to remove an invalid admin capability.

## Code

400 Bad Request

`NoSuchCap`

## Description

User does not possess specified capability.

Code

404 Not Found

## Quotas

The Admin Operations API enables you to set quotas on users and on bucket owned by users. See [Quota Management](#) for additional details. Quotas include the maximum number of objects in a bucket and the maximum storage size in megabytes.

To view quotas, the user must have a `users=read` capability. To set, modify or disable a quota, the user must have `users=write` capability. See the [Admin Guide](#) for details.

Valid parameters for quotas include:

- **Bucket:** The `bucket` option allows you to specify a quota for buckets owned by a user.
- **Maximum Objects:** The `max-objects` setting allows you to specify the maximum number of objects. A negative value disables this setting.
- **Maximum Size:** The `max-size` option allows you to specify a quota for the maximum number of bytes. The `max-size-kb` option allows you to specify it in KiB. A negative value disables this setting.
- **Quota Type:** The `quota-type` option sets the scope for the quota. The options are `bucket` and `user`.
- **Enable/Disable Quota:** The `enabled` option specifies whether the quota should be enabled. The value should be either 'True' or 'False'.

## Get User Quota

To get a quota, the user must have `users` capability set with `read` permission.

```
1. GET /admin/user?quota&uid=<uid>&quota-type=user
```

## Set User Quota

To set a quota, the user must have `users` capability set with `write` permission.

```
1. PUT /admin/user?quota&uid=<uid>&quota-type=user
```

The content must include a JSON representation of the quota settings as encoded in the corresponding read operation.

## Get Bucket Quota

To get a quota, the user must have `users` capability set with `read` permission.

```
1. GET /admin/user?quota&uid=<uid>&quota-type=bucket
```

## Set Bucket Quota

To set a quota, the user must have `users` capability set with `write` permission.

```
1. PUT /admin/user?quota&uid=<uid>&quota-type=bucket
```

The content must include a JSON representation of the quota settings as encoded in the corresponding read operation.

## Set Quota for an Individual Bucket

To set a quota, the user must have `buckets` capability set with `write` permission.

```
1. PUT /admin/bucket?quota&uid=<uid>&bucket=<bucket-name>&quota
```

The content must include a JSON representation of the quota settings as mentioned in Set Bucket Quota section above.

## Standard Error Responses

`AccessDenied`

Description

Access denied.

Code

403 Forbidden

`InternalServerError`

Description

Internal server error.

Code

500 Internal Server Error

`NoSuchUser`

## Description

User does not exist.

## Code

404 Not Found

NoSuchBucket

## Description

Bucket does not exist.

## Code

404 Not Found

NoSuchKey

## Description

No such access key.

## Code

404 Not Found

# Binding libraries

Golang

- [IrekFasikhov/go-rgwadmin](#)
- [QuentinPerez/go-radosgw](#)

Java

- [twonote/radosgw-admin4j](#)

Python

- [UMIACS/rgwadmin](#)
- [valerytschopp/python-radosgw-admin](#)

# NFS

---

New in version Jewel.

Ceph Object Gateway namespaces can now be exported over file-based access protocols such as NFSv3 and NFSv4, alongside traditional HTTP access protocols (S3 and Swift).

In particular, the Ceph Object Gateway can now be configured to provide file-based access when embedded in the NFS-Ganesha NFS server.

## librgw

---

The librgw.so shared library (Unix) provides a loadable interface to Ceph Object Gateway services, and instantiates a full Ceph Object Gateway instance on initialization.

In turn, librgw.so exports rgw\_file, a stateful API for file-oriented access to RGW buckets and objects. The API is general, but its design is strongly influenced by the File System Abstraction Layer (FSAL) API of NFS-Ganesha, for which it has been primarily designed.

A set of Python bindings is also provided.

## Namespace Conventions

---

The implementation conforms to Amazon Web Services (AWS) hierarchical namespace conventions which map UNIX-style path names onto S3 buckets and objects.

The top level of the attached namespace consists of S3 buckets, represented as NFS directories. Files and directories subordinate to buckets are each represented as objects, following S3 prefix and delimiter conventions, with ‘/’ being the only supported path delimiter [1](#).

For example, if an NFS client has mounted an RGW namespace at “/nfs”, then a file “/nfs/mybucket/www/index.html” in the NFS namespace corresponds to an RGW object “www/index.html” in a bucket/container “mybucket.”

Although it is generally invisible to clients, the NFS namespace is assembled through concatenation of the corresponding paths implied by the objects in the namespace. Leaf objects, whether files or directories, will always be materialized in an RGW object of the corresponding key name, “<name>” if a file, “<name>/” if a directory. Non-leaf directories (e.g., “www” above) might only be implied by their appearance in the names of one or more leaf objects. Directories created within NFS or directly operated on by an NFS client (e.g., via an attribute-setting operation such as chown or chmod) always have a leaf object representation used to store materialized attributes such as Unix

ownership and permissions.

## Supported Operations

---

The RGW NFS interface supports most operations on files and directories, with the following restrictions:

- Links, including symlinks, are not supported
- NFS ACLs are not supported
  - Unix user and group ownership and permissions are supported
- Directories may not be moved/renamed
  - files may be moved between directories
- Only full, sequential *write* i/o is supported
  - i.e., write operations are constrained to be **uploads**
  - many typical i/o operations such as editing files in place will necessarily fail as they perform non-sequential stores
  - some file utilities *apparently* writing sequentially (e.g., some versions of GNU tar) may fail due to infrequent non-sequential stores
  - When mounting via NFS, sequential application i/o can generally be constrained to be written sequentially to the NFS server via a synchronous mount option (e.g. `-osync` in Linux)
  - NFS clients which cannot mount synchronously (e.g., MS Windows) will not be able to upload files

## Security

---

The RGW NFS interface provides a hybrid security model with the following characteristics:

- NFS protocol security is provided by the NFS-Ganesha server, as negotiated by the NFS server and clients
  - e.g., clients can be trusted (AUTH\_SYS), or required to present Kerberos user credentials (RPCSEC\_GSS)
  - RPCSEC\_GSS wire security can be integrity only (krb5i) or integrity and privacy (encryption, krb5p)
  - various NFS-specific security and permission rules are available

- e.g., root-squashing
- a set of RGW/S3 security credentials (unknown to NFS) is associated with each RGW NFS mount (i.e., NFS-Ganesha EXPORT)
  - all RGW object operations performed via the NFS server will be performed by the RGW user associated with the credentials stored in the export being accessed (currently only RGW and RGW LDAP credentials are supported)
  - additional RGW authentication types such as Keystone are not currently supported

## Configuring an NFS-Ganesha Instance

Each NFS RGW instance is an NFS-Ganesha server instance *embedding* a full Ceph RGW instance.

Therefore, the RGW NFS configuration includes Ceph and Ceph Object Gateway-specific configuration in a local ceph.conf, as well as NFS-Ganesha-specific configuration in the NFS-Ganesha config file, ganesha.conf.

### ceph.conf

Required ceph.conf configuration for RGW NFS includes:

- valid [client.radosgw.{instance-name}] section
- valid values for minimal instance configuration, in particular, an installed and correct `keyring`

Other config variables are optional, front-end-specific and front-end selection variables (e.g., `rgw_data` and `rgw_frontends`) are optional and in some cases ignored.

A small number of config variables (e.g., `rgw_nfs_namespace_expire_secs`) are unique to RGW NFS.

### ganesha.conf

A strictly minimal ganesha.conf for use with RGW NFS includes one EXPORT block with embedded FSAL block of type RGW:

```

1. EXPORT
2. {
3.   Export_ID={numeric-id};
4.   Path = "/";
5.   Pseudo = "/";
6.   Access_Type = RW;
7.   SecType = "sys";
8.   NFS_Protocols = 4;

```

```

9.     Transport_Protocols = TCP;
10.
11.    # optional, permit unsquashed access by client "root" user
12.    #Squash = No_Root_Squash;
13.
14.    FSAL {
15.        Name = RGW;
16.        User_Id = {s3-user-id};
17.        Access_Key_Id ="{s3-access-key}";
18.        Secret_Access_Key = "{s3-secret}";
19.    }
20. }
```

`Export_ID` must have an integer value, e.g., "77"

`Path` (for RGW) should be "/"

`Pseudo` defines an NFSv4 pseudo root name (NFSv4 only)

`SecType = sys;` allows clients to attach without Kerberos authentication

`Squash = No_Root_Squash;` enables the client root user to override permissions (Unix convention). When root-squashing is enabled, operations attempted by the root user are performed as if by the local "nobody" (and "nogroup") user on the NFS-Ganesha server

The RGW FSAL additionally supports RGW-specific configuration variables in the RGW config section:

```

1. RGW {
2.     cluster = "{cluster name, default 'ceph'}";
3.     name = "client.rgw.{instance-name}";
4.     ceph_conf = "/opt/ceph-rgw/etc/ceph/ceph.conf";
5.     init_args = "-d --debug-rgw=16";
6. }
```

`cluster` sets a Ceph cluster name (must match the cluster being exported)

`name` sets an RGW instance name (must match the cluster being exported)

`ceph_conf` gives a path to a non-default ceph.conf file to use

## Other useful NFS-Ganesha configuration:

Any EXPORT block which should support NFSv3 should include version 3 in the NFS\_Protocols setting. Additionally, NFSv3 is the last major version to support the UDP transport. To enable UDP, include it in the Transport\_Protocols setting. For example:

```

1. EXPORT {
2. ...
3.     NFS_Protocols = 3,4;
```

```

4.     Transport_Protocols = UDP, TCP;
5. ...
6. }
```

One important family of options pertains to interaction with the Linux idmapping service, which is used to normalize user and group names across systems. Details of idmapper integration are not provided here.

With Linux NFS clients, NFS-Ganesha can be configured to accept client-supplied numeric user and group identifiers with NFSv4, which by default stringifies these—this may be useful in small setups and for experimentation:

```

1. NFSV4 {
2.     Allow_Numeric_Owners = true;
3.     Only_Numeric_Owners = true;
4. }
```

## Troubleshooting

NFS-Ganesha configuration problems are usually debugged by running the server with debugging options, controlled by the LOG config section.

NFS-Ganesha log messages are grouped into various components, logging can be enabled separately for each component. Valid values for component logging include:

```

1. *FATAL* critical errors only
2. *WARN* unusual condition
3. *DEBUG* mildly verbose trace output
4. *FULL_DEBUG* verbose trace output
```

Example:

```

1. LOG {
2.
3.     Components {
4.         MEMLEAKS = FATAL;
5.         FSAL = FATAL;
6.         NFSPROTO = FATAL;
7.         NFS_V4 = FATAL;
8.         EXPORT = FATAL;
9.         FILEHANDLE = FATAL;
10.        DISPATCH = FATAL;
11.        CACHE_INODE = FATAL;
12.        CACHE_INODE_LRU = FATAL;
13.        HASHTABLE = FATAL;
14.        HASHTABLE_CACHE = FATAL;
15.        DUPREQ = FATAL;
16.        INIT = DEBUG;
17.        MAIN = DEBUG;
```

```

18.         IDMAPPER = FATAL;
19.         NFS_READDIR = FATAL;
20.         NFS_V4_LOCK = FATAL;
21.         CONFIG = FATAL;
22.         CLIENTID = FATAL;
23.         SESSIONS = FATAL;
24.         PNFS = FATAL;
25.         RW_LOCK = FATAL;
26.         NLM = FATAL;
27.         RPC = FATAL;
28.         NFS_CB = FATAL;
29.         THREAD = FATAL;
30.         NFS_V4_ACL = FATAL;
31.         STATE = FATAL;
32.         FSAL_UP = FATAL;
33.         DBUS = FATAL;
34.     }
35.     # optional: redirect log output
36.     # Facility {
37.     #         name = FILE;
38.     #         destination = "/tmp/ganesha-rgw.log";
39.     #         enable = active;
40.     }
41. }
```

## Running Multiple NFS Gateways

Each NFS-Ganesha instance acts as a full gateway endpoint, with the limitation that currently an NFS-Ganesha instance cannot be configured to export HTTP services. As with ordinary gateway instances, any number of NFS-Ganesha instances can be started, exporting the same or different resources from the cluster. This enables the clustering of NFS-Ganesha instances. However, this does not imply high availability.

When regular gateway instances and NFS-Ganesha instances overlap the same data resources, they will be accessible from both the standard S3 API and through the NFS-Ganesha instance as exported. You can co-locate the NFS-Ganesha instance with a Ceph Object Gateway instance on the same host.

## RGW vs RGW NFS

Exporting an NFS namespace and other RGW namespaces (e.g., S3 or Swift via the Civetweb HTTP front-end) from the same program instance is currently not supported.

When adding objects and buckets outside of NFS, those objects will appear in the NFS namespace in the time set by `rgw_nfs_namespace_expire_secs`, which defaults to 300 seconds (5 minutes). Override the default value for `rgw_nfs_namespace_expire_secs` in the Ceph configuration file to change the refresh rate.

If exporting Swift containers that do not conform to valid S3 bucket naming

requirements, set `rgw_relaxed_s3_bucket_names` to true in the [client.radosgw] section of the Ceph configuration file. For example, if a Swift container name contains underscores, it is not a valid S3 bucket name and will be rejected unless `rgw_relaxed_s3_bucket_names` is set to true.

## Configuring NFSv4 clients

To access the namespace, mount the configured NFS-Ganesha export(s) into desired locations in the local POSIX namespace. As noted, this implementation has a few unique restrictions:

- NFS 4.1 and higher protocol flavors are preferred
  - NFSv4 OPEN and CLOSE operations are used to track upload transactions
- To upload data successfully, clients must preserve write ordering
  - on Linux and many Unix NFS clients, use the `-osync` mount option

Conventions for mounting NFS resources are platform-specific. The following conventions work on Linux and some Unix platforms:

From the command line:

```
1. mount -t nfs -o nfsvers=4.1,noauto,soft,sync,proto=tcp <ganesha-host-name>:/ <mount-point>
```

In /etc/fstab:

```
1. <ganesha-host-name>:/ <mount-point> nfs noauto,soft,nfsvers=4.1,sync,proto=tcp 0 0
```

Specify the NFS-Ganesha host name and the path to the mount point on the client.

## Configuring NFSv3 Clients

Linux clients can be configured to mount with NFSv3 by supplying `nfsvers=3` and `noacl` as mount options. To use UDP as the transport, add `proto=udp` to the mount options. However, TCP is the preferred transport:

```
1. <ganesha-host-name>:/ <mount-point> nfs noauto,noacl,soft,nfsvers=3,sync,proto=tcp 0 0
```

Configure the NFS Ganesha EXPORT block Protocols setting with version 3 and the Transports setting with UDP if the mount will use version 3 with UDP.

## NFSv3 Semantics

Since NFSv3 does not communicate client OPEN and CLOSE operations to file servers, RGW

NFS cannot use these operations to mark the beginning and ending of file upload transactions. Instead, RGW NFS starts a new upload when the first write is sent to a file at offset 0, and finalizes the upload when no new writes to the file have been seen for a period of time, by default, 10 seconds. To change this timeout, set an alternate value for `rgw_nfs_write_completion_interval_s` in the RGW section(s) of the Ceph configuration file.

## References

---

1

<http://docs.aws.amazon.com/AmazonS3/latest/dev/ListingKeysHierarchy.html>

# Integrating with OpenStack Keystone

It is possible to integrate the Ceph Object Gateway with Keystone, the OpenStack identity service. This sets up the gateway to accept Keystone as the users authority. A user that Keystone authorizes to access the gateway will also be automatically created on the Ceph Object Gateway (if didn't exist beforehand). A token that Keystone validates will be considered as valid by the gateway.

The following configuration options are available for Keystone integration:

1. [client.radosgw.gateway]
2. rgw keystone api version = {keystone api version}
3. rgw keystone url = {keystone server url:keystone server admin port}
4. rgw keystone admin token = {keystone admin token}
5. rgw keystone admin token path = {path to keystone admin token} #preferred
6. rgw keystone accepted roles = {accepted user roles}
7. rgw keystone token cache size = {number of tokens to cache}
8. rgw keystone implicit tenants = {true for private tenant for each new user}

It is also possible to configure a Keystone service tenant, user & password for Keystone (for v2.0 version of the OpenStack Identity API), similar to the way OpenStack services tend to be configured, this avoids the need for setting the shared secret `rgw keystone admin token` in the configuration file, which is recommended to be disabled in production environments. The service tenant credentials should have admin privileges, for more details refer the [OpenStack Keystone documentation](#), which explains the process in detail. The requisite configuration options for are:

1. rgw keystone admin user = {keystone service tenant user name}
2. rgw keystone admin password = {keystone service tenant user password}
3. rgw keystone admin password = {keystone service tenant user password path} # preferred
4. rgw keystone admin tenant = {keystone service tenant name}

A Ceph Object Gateway user is mapped into a Keystone `tenant`. A Keystone user has different roles assigned to it on possibly more than a single tenant. When the Ceph Object Gateway gets the ticket, it looks at the tenant, and the user roles that are assigned to that ticket, and accepts/rejects the request according to the `rgw keystone accepted roles` configurable.

For a v3 version of the OpenStack Identity API you should replace `rgw keystone admin tenant` with:

1. rgw keystone admin domain = {keystone admin domain name}
2. rgw keystone admin project = {keystone admin project name}

For compatibility with previous versions of ceph, it is also possible to set `rgw keystone implicit tenants` to either `s3` or `swift`. This has the effect of splitting the

identity space such that the indicated protocol will only use implicit tenants, and the other protocol will never use implicit tenants. Some older versions of ceph only supported implicit tenants with swift.

## Ocata (and later)

Keystone itself needs to be configured to point to the Ceph Object Gateway as an object-storage endpoint:

```
1. openstack service create --name=swift \
2.                               --description="Swift Service" \
3.                               object-store
4. +-----+-----+
5. | Field      | Value          |
6. +-----+-----+
7. | description | Swift Service |
8. | enabled     | True           |
9. | id          | 37c4c0e79571404cb4644201a4a6e5ee |
10. | name        | swift          |
11. | type        | object-store   |
12. +-----+-----+
13.
14. openstack endpoint create --region RegionOne \
15.     --publicurl "http://radosgw.example.com:8080/swift/v1" \
16.     --adminurl  "http://radosgw.example.com:8080/swift/v1" \
17.     --internalurl "http://radosgw.example.com:8080/swift/v1" \
18.     swift
19. +-----+-----+
20. | Field      | Value          |
21. +-----+-----+
22. | adminurl   | http://radosgw.example.com:8080/swift/v1 |
23. | id          | e4249d2b60e44743a67b5e5b38c18dd3 |
24. | internalurl | http://radosgw.example.com:8080/swift/v1 |
25. | publicurl   | http://radosgw.example.com:8080/swift/v1 |
26. | region      | RegionOne      |
27. | service_id  | 37c4c0e79571404cb4644201a4a6e5ee |
28. | service_name| swift          |
29. | service_type| object-store   |
30. +-----+-----+
31.
32. $ openstack endpoint show object-store
33. +-----+-----+
34. | Field      | Value          |
35. +-----+-----+
36. | adminurl   | http://radosgw.example.com:8080/swift/v1 |
37. | enabled     | True           |
38. | id          | e4249d2b60e44743a67b5e5b38c18dd3 |
39. | internalurl | http://radosgw.example.com:8080/swift/v1 |
40. | publicurl   | http://radosgw.example.com:8080/swift/v1 |
41. | region      | RegionOne      |
42. | service_id  | 37c4c0e79571404cb4644201a4a6e5ee |
```

```

43. | service_name | swift
44. | service_type | object-store
45. +-----+-----+

```

## Note

If your radosgw `ceph.conf` sets the configuration option `rgw swift account in url = true`, your `object-store` endpoint URLs must be set to include the suffix `/v1/AUTH_%(tenant_id)s` (instead of just `/v1`).

The Keystone URL is the Keystone admin RESTful API URL. The admin token is the token that is configured internally in Keystone for admin requests.

OpenStack Keystone may be terminated with a self signed ssl certificate, in order for radosgw to interact with Keystone in such a case, you could either install Keystone's ssl certificate in the node running radosgw. Alternatively radosgw could be made to not verify the ssl certificate at all (similar to OpenStack clients with a `--insecure` switch) by setting the value of the configurable `rgw keystone verify ssl` to false.

## Cross Project(Tenant) Access

In order to let a project (earlier called a 'tenant') access buckets belonging to a different project, the following config option needs to be enabled:

```
1. rgw swift account in url = true
```

The Keystone object-store endpoint must accordingly be configured to include the `AUTH_%(project_id)s` suffix:

```

1. openstack endpoint create --region RegionOne \
2.   --publicurl "http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s" \
3.   --adminurl "http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s" \
4.   --internalurl "http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s" \
5.   swift
6. +-----+-----+
7. | Field      | Value
8. +-----+-----+
9. | adminurl    | http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s |
10. | id          | e4249d2b60e44743a67b5e5b38c18dd3 |
11. | internalurl | http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s |
12. | publicurl   | http://radosgw.example.com:8080/swift/v1/AUTH_%(project_id)s |
13. | region       | RegionOne |
14. | service_id   | 37c4c0e79571404cb4644201a4a6e5ee |
15. | service_name | swift |
16. | service_type | object-store |
17. +-----+-----+

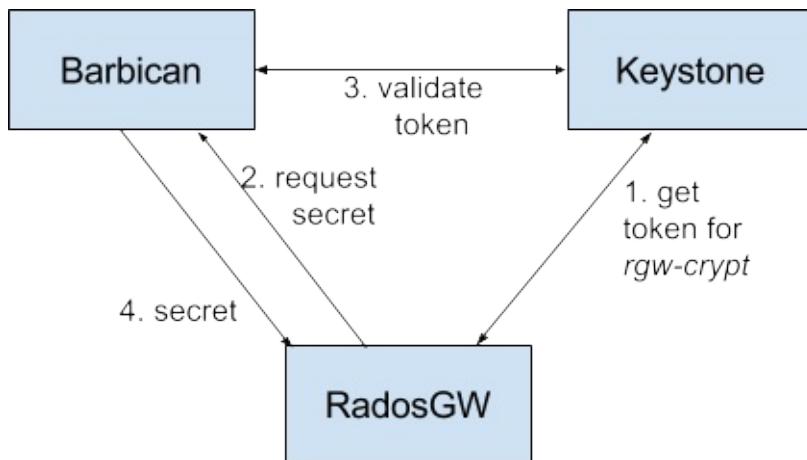
```

## Keystone integration with the S3 API

It is possible to use Keystone for authentication even when using the S3 API (with AWS-like access and secret keys), if the `rgw s3 auth use keystone` option is set. For details, see [Authentication and ACLs](#).

# OpenStack Barbican Integration

OpenStack [Barbican](#) can be used as a secure key management service for [Server-Side Encryption](#).



1. [Configure Keystone](#)
2. [Create a Keystone user](#)
3. [Configure the Ceph Object Gateway](#)
4. [Create a key in Barbican](#)

## Configure Keystone

Barbican depends on Keystone for authorization and access control of its keys.

See [OpenStack Keystone Integration](#).

## Create a Keystone user

Create a new user that will be used by the Ceph Object Gateway to retrieve keys.

For example:

```
1. user = rgwcrypt-user
2. pass = rgwcrypt-password
3. tenant = rgwcrypt
```

See OpenStack documentation for [Manage projects, users, and roles](#).

## Create a key in Barbican

See Barbican documentation for [How to Create a Secret](#). Requests to Barbican must include a valid Keystone token in the `X-Auth-Token` header.

## Note

Server-side encryption keys must be 256-bit long and base64 encoded.

Example request:

```

1. POST /v1/secrets HTTP/1.1
2. Host: barbican.example.com:9311
3. Accept: */*
4. Content-Type: application/json
5. X-Auth-Token: 7f7d588dd29b44df983bc961a6b73a10
6. Content-Length: 299
7.
8. {
9.     "name": "my-key",
10.    "expiration": "2016-12-28T19:14:44.180394",
11.    "algorithm": "aes",
12.    "bit_length": 256,
13.    "mode": "cbc",
14.    "payload": "6b+W0Z1T3cqZMxgThRcXAQBrS5mXKdDUphvxptl9/4=",
15.    "payload_content_type": "application/octet-stream",
16.    "payload_content_encoding": "base64"
17. }
```

Response:

```
1. {"secret_ref": "http://barbican.example.com:9311/v1/secrets/d1e7ef3b-f841-4b7c-90b2-b7d90ca2d723"}
```

In the response, `d1e7ef3b-f841-4b7c-90b2-b7d90ca2d723` is the key id that can be used in any SSE-KMS request.

This newly created key is not accessible by user `rgwcrypt-user`. This privilege must be added with an ACL. See [How to Set/Replace ACL](#) for more details.

Example request (assuming that the Keystone id of `rgwcrypt-user` is `906aa90bd8a946c89cdff80d0869460f`):

```

1. PUT /v1/secrets/d1e7ef3b-f841-4b7c-90b2-b7d90ca2d723/acl HTTP/1.1
2. Host: barbican.example.com:9311
3. Accept: */*
4. Content-Type: application/json
5. X-Auth-Token: 7f7d588dd29b44df983bc961a6b73a10
6. Content-Length: 101
7.
8. {
9.     "read": {
10.        "users": [
11.            "906aa90bd8a946c89cdff80d0869460f"
12.        ],
13.        "project-access": true
14.    }
15. }
```

```
12.      }
13. }
```

Response:

```
1. {"acl_ref": "http://barbican.example.com:9311/v1/secrets/d1e7ef3b-f841-4b7c-90b2-b7d90ca2d723/acl"}
```

## Configure the Ceph Object Gateway

Edit the Ceph configuration file to enable Barbican as a KMS and add information about the Barbican server and Keystone user:

```
1. rgw crypt s3 kms backend = barbican
2. rgw barbican url = http://barbican.example.com:9311
3. rgw keystone barbican user = rgwcrypt-user
4. rgw keystone barbican password = rgwcrypt-password
```

When using Keystone API version 2:

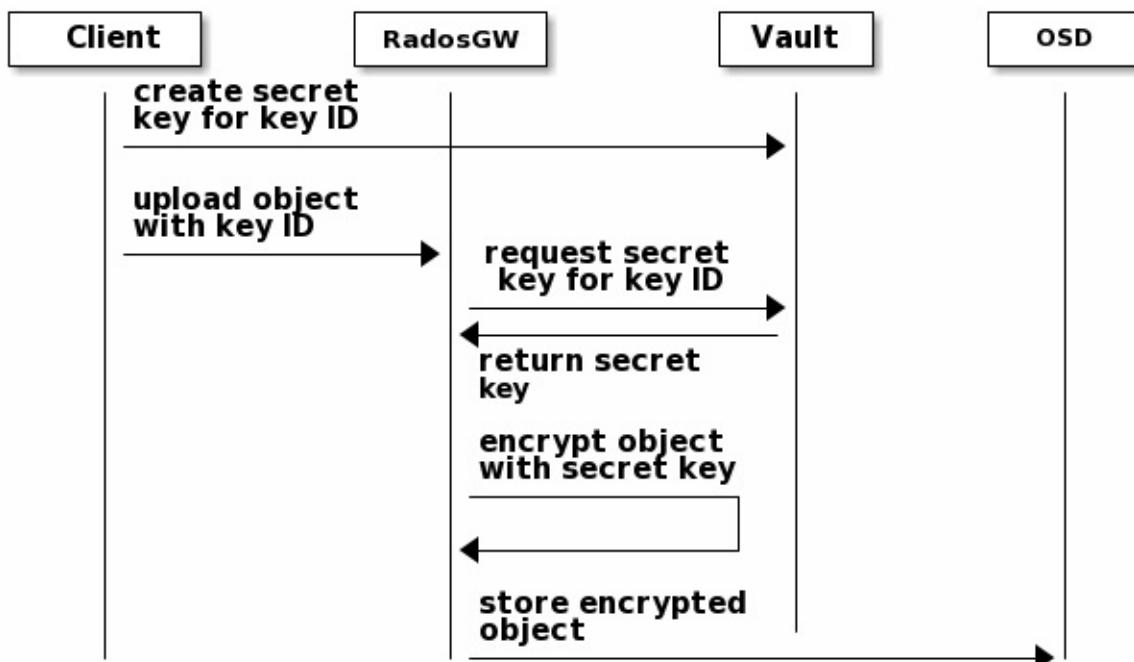
```
1. rgw keystone barbican tenant = rgwcrypt
```

When using API version 3:

```
1. rgw keystone barbican project
2. rgw keystone barbican domain
```

# HashiCorp Vault Integration

HashiCorp [Vault](#) can be used as a secure key management service for [Server-Side Encryption \(SSE-KMS\)](#).



1. Vault secrets engines
  2. Vault authentication
  3. Vault namespaces
  4. Create a key in Vault
  5. Configure the Ceph Object Gateway
  6. Upload object

Some examples below use the Vault command line utility to interact with Vault. You may need to set the following environment variable with the correct address of your Vault server to use this utility:

```
1. export VAULT_ADDR='http://vault-server:8200'
```

## Vault secrets engines

Vault provides several secrets engines, which can store, generate, and encrypt data.

Currently, the Object Gateway supports:

- [KV secrets engine version 2](#)
- [Transit engine](#)

## KV secrets engine

The KV secrets engine is used to store arbitrary key/value secrets in Vault. To enable the KV engine version 2 in Vault, use the following command:

```
1. vault secrets enable -path secret kv-v2
```

The Object Gateway can be configured to use the KV engine version 2 with the following setting:

```
1. rgw crypt vault secret engine = kv
```

## Transit secrets engine

The transit engine handles cryptographic functions on data in-transit. To enable it in Vault, use the following command:

```
1. vault secrets enable transit
```

The Object Gateway can be configured to use the transit engine with the following setting:

```
1. rgw crypt vault secret engine = transit
```

## Vault authentication

Vault supports several authentication mechanisms. Currently, the Object Gateway can be configured to authenticate to Vault using the [Token authentication method](#) or a [Vault agent](#).

### Token authentication

#### Note

Token authentication is not recommended for production environments.

The token authentication method expects a Vault token to be present in a plaintext file. The Object Gateway can be configured to use token authentication with the following settings:

```

1. rgw crypt vault auth = token
2. rgw crypt vault token file = /etc/ceph/vault.token
3. rgw crypt vault addr = http://vault-server:8200

```

For security reasons, the token file must be readable by the Object Gateway only. Also, the Object Gateway should be given a Vault token with a restricted policy that allows it to fetch keyrings from a specific path only. Such a policy can be created in Vault using the command line utility as in the following examples:

```

1. vault policy write rgw-kv-policy -<<EOF
2.   path "secret/data/*" {
3.     capabilities = ["read"]
4.   }
5. EOF
6.
7. vault policy write rgw-transit-policy -<<EOF
8.   path "transit/export/encryption-key/*" {
9.     capabilities = ["read"]
10.  }
11. EOF

```

Once the policy is created, a token can be generated by a Vault administrator:

```
1. vault token create -policy=rgw-kv-policy
```

Sample output:

| Key               | Value                       |
|-------------------|-----------------------------|
| ---               | -----                       |
| token             | s.72KuPujbc0650dWB71po0mIq  |
| token_accessor    | jv95ZYBUFv6Ss84x7SCSy6lZ    |
| token_duration    | 768h                        |
| token_renewable   | true                        |
| token_policies    | ["default" "rgw-kv-policy"] |
| identity_policies | []                          |
| policies          | ["default" "rgw-kv-policy"] |

The actual token, displayed in the `value` column of the first line of the output, must be saved in a file as plaintext.

## Vault agent

The Vault agent is a client daemon that provides authentication to Vault and manages token renewal and caching. It typically runs on the same host as the Object Gateway. With a Vault agent, it is possible to use other Vault authentication mechanism such as AppRole, AWS, Certs, JWT, and Azure.

The Object Gateway can be configured to use a Vault agent with the following settings:

```
1. rgw crypt vault auth = agent
2. rgw crypt vault addr = http://localhost:8100
```

## Vault namespaces

In the Enterprise version, Vault supports the concept of [namespaces](#), which allows centralized management for teams within an organization while ensuring that those teams operate within isolated environments known as tenants.

The Object Gateway can be configured to access Vault within a particular namespace using the following configuration setting:

```
1. rgw crypt vault namespace = tenant1
```

## Create a key in Vault

### Note

Keys for server-side encryption must be 256-bit long and base-64 encoded.

## Using the KV engine

A key for server-side encryption can be created in the KV version 2 engine using the command line utility, as in the following example:

```
1. vault kv put secret/myproject/mybucketkey key=$(openssl rand -base64 32)
```

### Sample output:

```
1. ===== Metadata =====
2. Key          Value
3. ---
4. created_time 2019-08-29T17:01:09.095824999Z
5. deletion_time n/a
6. destroyed    false
7. version      1
```

Note that in the KV secrets engine, secrets are stored as key-value pairs, and the Gateway expects the key name to be `key`, i.e. the secret must be in the form `key=<secret key>`.

## Using the Transit engine

Keys created with the Transit engine must be exportable in order to be used for server-side encryption with the Object Gateway. An exportable key can be created with

the command line utility as follows:

```
1. vault write -f transit/keys/mybucketkey exportable=true
```

The command above creates a keyring, which contains a key of type `aes256-gcm96` by default. To verify that the key was correctly created, use the following command:

```
1. vault read transit/export/encryption-key/mybucketkey/1
```

Sample output:

| Key  | Value                                                           |
|------|-----------------------------------------------------------------|
| ---  | -----                                                           |
| keys | <code>map[1:-gbTI9lNpqv/V/21DcmH2Nq1xKn6FPDWarCmFM2aNQ=]</code> |
| name | mybucketkey                                                     |
| type | aes256-gcm96                                                    |

Note that in order to read the key created with the Transit engine, the full path must be provided including the key version.

## Configure the Ceph Object Gateway

Edit the Ceph configuration file to enable Vault as a KMS backend for server-side encryption:

```
1. rgw crypt s3 kms backend = vault
```

Choose the Vault authentication method, e.g.:

```
1. rgw crypt vault auth = token
2. rgw crypt vault token file = /etc/ceph/vault.token
3. rgw crypt vault addr = http://vault-server:8200
```

Or:

```
1. rgw crypt vault auth = agent
2. rgw crypt vault addr = http://localhost:8100
```

Choose the secrets engine:

```
1. rgw crypt vault secret engine = kv
```

Or:

```
1. rgw crypt vault secret engine = transit
```

Optionally, set the Vault namespace where encryption keys will be fetched from:

```
1. rgw crypt vault namespace = tenant1
```

Finally, the URLs where the Gateway will retrieve encryption keys from Vault can be restricted by setting a path prefix. For instance, the Gateway can be restricted to fetch KV keys as follows:

```
1. rgw crypt vault prefix = /v1/secret/data
```

Or, in the case of exportable transit keys:

```
1. rgw crypt vault prefix = /v1/transit/export/encryption-key
```

In the example above, the Gateway would only fetch transit encryption keys under `http://vault-server:8200/v1/transit/export/encryption-key`.

## Upload object

When uploading an object to the Gateway, provide the SSE key ID in the request. As an example, for the kv engine, using the AWS command-line client:

```
aws --endpoint=http://radosgw:8000 s3 cp plaintext.txt s3://mybucket/encrypted.txt --sse=aws:kms --sse-kms-key-id myproject/mybucketkey
```

As an example, for the transit engine, using the AWS command-line client:

```
aws --endpoint=http://radosgw:8000 s3 cp plaintext.txt s3://mybucket/encrypted.txt --sse=aws:kms --sse-kms-key-id mybucketkey/1
```

The Object Gateway will fetch the key from Vault, encrypt the object and store it in the bucket. Any request to download the object will make the Gateway automatically retrieve the correspondent key from Vault and decrypt the object.

Note that the secret will be fetched from Vault using a URL constructed by concatenating the base address (`rgw crypt vault addr`), the (optional) URL prefix (`rgw crypt vault prefix`), and finally the key ID.

In the kv engine example above, the Gateway would fetch the secret from:

```
1. http://vaultserver:8200/v1/secret/data/myproject/mybucketkey
```

In the transit engine example above, the Gateway would fetch the secret from:

```
1. http://vaultserver:8200/v1/transit/export/encryption-key/mybucketkey/1
```



# Open Policy Agent Integration

Open Policy Agent (OPA) is a lightweight general-purpose policy engine that can be co-located with a service. OPA can be integrated as a sidecar, host-level daemon, or library.

Services can offload policy decisions to OPA by executing queries. Hence, policy enforcement can be decoupled from policy decisions.

## Configure OPA

To configure OPA, load custom policies into OPA that control the resources users are allowed to access. Relevant data or context can also be loaded into OPA to make decisions.

Policies and data can be loaded into OPA in the following ways::

- OPA's RESTful APIs
- OPA's *bundle* feature that downloads policies and data from remote HTTP servers
- Filesystem

## Configure the Ceph Object Gateway

The following configuration options are available for OPA integration:

1. rgw use opa authz = {use opa server to authorize client requests}
2. rgw opa url = {opa server url:opa server port}
3. rgw opa token = {opa bearer token}
4. rgw opa verify ssl = {verify opa server ssl certificate}

## How does the RGW-OPA integration work

After a user is authenticated, OPA can be used to check if the user is authorized to perform the given action on the resource. OPA responds with an allow or deny decision which is sent back to the RGW which enforces the decision.

Example request:

1. POST /v1/data/ceph/authz HTTP/1.1
2. Host: opa.example.com:8181
3. Content-Type: application/json
- 4.
5. {

```
6.     "input": {
7.         "method": "GET",
8.         "subuser": "subuser",
9.         "user_info": {
10.             "user_id": "john",
11.             "display_name": "John"
12.         },
13.         "bucket_info": {
14.             "bucket": {
15.                 "name": "Testbucket",
16.                 "bucket_id": "testbucket"
17.             },
18.             "owner": "john"
19.         }
20.     }
21. }
```

Response:

```
1. {"result": true}
```

The above is a sample request sent to OPA which contains information about the user, resource and the action to be performed on the resource. Based on the policies and data loaded into OPA, it will verify whether the request should be allowed or denied. In the sample request, RGW makes a POST request to the endpoint `/v1/data/ceph/authz`, where `ceph` is the package name and `authz` is the rule name.

# RGW Multi-tenancy

New in version Jewel.

The multi-tenancy feature allows to use buckets and users of the same name simultaneously by segregating them under so-called `tenants`. This may be useful, for instance, to permit users of Swift API to create buckets with easily conflicting names such as "test" or "trove".

From the Jewel release onward, each user and bucket lies under a tenant. For compatibility, a "legacy" tenant with an empty name is provided. Whenever a bucket is referred without an explicit tenant, an implicit tenant is used, taken from the user performing the operation. Since the pre-existing users are under the legacy tenant, they continue to create and access buckets as before. The layout of objects in RADOS is extended in a compatible way, ensuring a smooth upgrade to Jewel.

## Administering Users With Explicit Tenants

Tenants as such do not have any operations on them. They appear and disappear as needed, when users are administered. In order to create, modify, and remove users with explicit tenants, either an additional option `-tenant` is supplied, or a syntax "`<tenant>$<user>`" is used in the parameters of the `radosgw-admin` command.

### Examples

Create a user `testx$tester` to be accessed with S3:

```
# radosgw-admin --tenant testx --uid tester --display-name "Test User" --access_key TESTER --secret test123
1. user create
```

Create a user `testx$tester` to be accessed with Swift:

```
# radosgw-admin --tenant testx --uid tester --display-name "Test User" --subuser tester:test --key-type swift
1. --access full user create
2. # radosgw-admin --subuser 'testx$tester:test' --key-type swift --secret test123
```

#### Note

The subuser with explicit tenant has to be quoted in the shell.

Tenant names may contain only alphanumeric characters and underscores.

## Accessing Buckets with Explicit Tenants

When a client application accesses buckets, it always operates with credentials of a

particular user. As mentioned above, every user belongs to a tenant. Therefore, every operation has an implicit tenant in its context, to be used if no tenant is specified explicitly. Thus a complete compatibility is maintained with previous releases, as long as the referred buckets and referring user belong to the same tenant. In other words, anything unusual occurs when accessing another tenant's buckets *only*.

Extensions employed to specify an explicit tenant differ according to the protocol and authentication system used.

## S3

In case of S3, a colon character is used to separate tenant and bucket. Thus a sample URL would be:

```
1. https://ep.host.dom:tenant:bucket
```

Here's a simple Python sample:

|                                               |                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre>1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 7. 7</pre> | <pre>1. from boto.s3.connection import S3Connection, OrdinaryCallingFormat 2. c = S3Connection( 3.     aws_access_key_id="TESTER", 4.     aws_secret_access_key="test123", 5.     host="ep.host.dom", 6.     calling_format = OrdinaryCallingFormat()) 7. bucket = c.get_bucket("test5b:testbucket")</pre> |
|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Note that it's not possible to supply an explicit tenant using a hostname. Hostnames cannot contain colons, or any other separators that are not already valid in bucket names. Using a period creates an ambiguous syntax. Therefore, the bucket-in-URL-path format has to be used.

Due to the fact that the native S3 API does not deal with multi-tenancy and radosgw's implementation does, things get a bit involved when dealing with signed URLs and public read ACLs.

- A **signed URL** does contain the `AWSAccessKeyId` query parameters, from which radosgw is able to discern the correct user and tenant owning the bucket. In other words, an application generating signed URLs should be able to take just the un-prefixed bucket name, and produce a signed URL that itself contains the bucket name without the tenant prefix. However, it is possible to include the prefix if you so choose.

Thus, accessing a signed URL of an object `bar` in a container `foo` belonging to the tenant `7188e165c0ae4424ac68ae2e89a05c50` would be possible either via `http://<host>:<port>/foo/bar?`

`AWSAccessKeyId=b200fb6634c547199e436a0f93c0c46e&Expires=1542890806&Signature=eok6CYQC%2FDwmQQmqvY5jTg6ehXU%3D`,

or via `http://<host>/7188e165c0ae4424ac68ae2e89a05c50:foo/bar?`  
`AWSAccessKeyId=b200fb6634c547199e436a0f93c0c46e&Expires=1542890806&Signature=eok6CYQC%2FDwmQQmqvY5jTg6ehXU%3D`, depending on whether or not the tenant prefix was passed in on signature generation.

- A bucket with a **public read ACL** is meant to be read by an HTTP client *without* including any query parameters that would allow radosgw to discern tenants. Thus, publicly readable objects must always be accessed using the bucket name with the tenant prefix.

Thus, if you set a public read ACL on an object `bar` in a container `foo` belonging to the tenant `7188e165c0ae4424ac68ae2e89a05c50`, you would need to access that object via the public URL `http://<host>/7188e165c0ae4424ac68ae2e89a05c50:foo/bar`.

## Swift with built-in authenticator

TBD – not in test\_multen.py yet

## Swift with Keystone

In the default configuration, although native Swift has inherent multi-tenancy, radosgw does not enable multi-tenancy for the Swift API. This is to ensure that a setup with legacy buckets – that is, buckets that were created before radosgw supported multitenancy –, those buckets retain their dual-API capability to be queried and modified using either S3 or Swift.

If you want to enable multitenancy for Swift, particularly if your users only ever authenticate against OpenStack Keystone, you should enable Keystone-based multitenancy with the following `ceph.conf` configuration option:

```
1. rgw keystone implicit_tenants = true
```

Once you enable this option, any newly connecting user (whether they are using the Swift API, or Keystone-authenticated S3) will prompt radosgw to create a user named `<tenant_id>$<tenant_id>`, where `<tenant_id>` is a Keystone tenant (project) UUID – for example, `7188e165c0ae4424ac68ae2e89a05c50$7188e165c0ae4424ac68ae2e89a05c50`.

Whenever that user then creates an Swift container, radosgw internally translates the given container name into `<tenant_id>/<container_name>`, such as `7188e165c0ae4424ac68ae2e89a05c50/foo`. This ensures that if there are two or more different tenants all creating a container named `foo`, radosgw is able to transparently discern them by their tenant prefix.

It is also possible to limit the effects of implicit tenants to only apply to swift or s3, by setting `rgw keystone implicit_tenants` to either `s3` or `swift`. This will likely primarily be of use to users who had previously used implicit tenants with older versions of ceph, where implicit tenants only applied to the swift protocol.

## Notes and known issues

Just to be clear, it is not possible to create buckets in other tenants at present. The owner of newly created bucket is extracted from authentication information.

# Compression

New in version Kraken.

The Ceph Object Gateway supports server-side compression of uploaded objects, using any of Ceph's existing compression plugins.

## Configuration

Compression can be enabled on a storage class in the Zone's placement target by providing the `--compression=<type>` option to the command `radosgw-admin zone placement modify`.

The compression `type` refers to the name of the compression plugin to use when writing new object data. Each compressed object remembers which plugin was used, so changing this setting does not hinder the ability to decompress existing objects, nor does it force existing objects to be recompressed.

This compression setting applies to all new objects uploaded to buckets using this placement target. Compression can be disabled by setting the `type` to an empty string or `none`.

For example:

```
1. $ radosgw-admin zone placement modify \
2.   --rgw-zone default \
3.   --placement-id default-placement \
4.   --storage-class STANDARD \
5.   --compression zlib
6. {
7. ...
8.   "placement_pools": [
9.     {
10.       "key": "default-placement",
11.       "val": {
12.         "index_pool": "default.rgw.buckets.index",
13.         "storage_classes": {
14.           "STANDARD": {
15.             "data_pool": "default.rgw.buckets.data",
16.             "compression_type": "zlib"
17.           }
18.         },
19.         "data_extra_pool": "default.rgw.buckets.non-ec",
20.         "index_type": 0,
21.       }
22.     }
23.   ],
24. ...
25. }
```

## Note

A `default` zone is created for you if you have not done any previous [Multisite Configuration](#).

## Statistics

While all existing commands and APIs continue to report object and bucket sizes based their uncompressed data, compression statistics for a given bucket are included in its `bucket stats` :

```
1. $ radosgw-admin bucket stats --bucket=<name>
2. {
3. ...
4.     "usage": {
5.         "rgw.main": {
6.             "size": 1075028,
7.             "size_actual": 1331200,
8.             "size_utilized": 592035,
9.             "size_kb": 1050,
10.            "size_kb_actual": 1300,
11.            "size_kb_utilized": 579,
12.            "num_objects": 104
13.        }
14.    },
15. ...
16. }
```

The `size_utilized` and `size_kb_utilized` fields represent the total size of compressed data, in bytes and kilobytes respectively.

# LDAP Authentication

New in version Jewel.

You can delegate the Ceph Object Gateway authentication to an LDAP server.

## How it works

The Ceph Object Gateway extracts the users LDAP credentials from a token. A search filter is constructed with the user name. The Ceph Object Gateway uses the configured service account to search the directory for a matching entry. If an entry is found, the Ceph Object Gateway attempts to bind to the found distinguished name with the password from the token. If the credentials are valid, the bind will succeed, and the Ceph Object Gateway will grant access and radosgw-user will be created with the provided username.

You can limit the allowed users by setting the base for the search to a specific organizational unit or by specifying a custom search filter, for example requiring specific group membership, custom object classes, or attributes.

The LDAP credentials must be available on the server to perform the LDAP authentication. Make sure to set the `rgw` log level low enough to hide the base-64-encoded credentials / access tokens.

## Requirements

- **LDAP or Active Directory:** A running LDAP instance accessible by the Ceph Object Gateway
- **Service account:** LDAP credentials to be used by the Ceph Object Gateway with search permissions
- **User account:** At least one user account in the LDAP directory
- **Do not overlap LDAP and local users:** You should not use the same user names for local users and for users being authenticated by using LDAP. The Ceph Object Gateway cannot distinguish them and it treats them as the same user.

## Sanity checks

Use the `ldapsearch` utility to verify the service account or the LDAP connection:

```
1. # ldapsearch -x -D "uid=ceph,ou=system,dc=example,dc=com" -W \
2. -H ldaps://example.com -b "ou=users,dc=example,dc=com" 'uid=' dn
```

## Note

Make sure to use the same LDAP parameters like in the Ceph configuration file to eliminate possible problems.

# Configuring the Ceph Object Gateway to use LDAP authentication

The following parameters in the Ceph configuration file are related to the LDAP authentication:

- `rgw_s3_auth_use_ldap` : Set this to `true` to enable S3 authentication with LDAP
- `rgw_ldap_uri` : Specifies the LDAP server to use. Make sure to use the `ldaps://<fqdn>:<port>` parameter to not transmit clear text credentials over the wire.
- `rgw_ldap_binddn` : The Distinguished Name (DN) of the service account used by the Ceph Object Gateway
- `rgw_ldap_secret` : Path to file containing credentials for `rgw_ldap_binddn`
- `rgw_ldap_searchdn` : Specifies the base in the directory information tree for searching users. This might be your users organizational unit or some more specific Organizational Unit (OU).
- `rgw_ldap_dnattr` : The attribute being used in the constructed search filter to match a username. Depending on your Directory Information Tree (DIT) this would probably be `uid` or `cn`. The generated filter string will be, e.g., `cn=some_username`.
- `rgw_ldap_searchfilter` : If not specified, the Ceph Object Gateway automatically constructs the search filter with the `rgw_ldap_dnattr` setting. Use this parameter to narrow the list of allowed users in very flexible ways. Consult the *Using a custom search filter to limit user access* section for details

## Using a custom search filter to limit user access

There are two ways to use the `rgw_search_filter` parameter:

### Specifying a partial filter to further limit the constructed search filter

An example for a partial filter:

```
1. "objectclass=inetorgperson"
```

The Ceph Object Gateway will generate the search filter as usual with the user name from the token and the value of `rgw_ldap_dnattr`. The constructed filter is then combined with the partial filter from the `rgw_search_filter` attribute. Depending on the user name and the settings the final search filter might become:

```
1. "(&(uid=hari)(objectclass=inetorgperson))"
```

So user `hari` will only be granted access if he is found in the LDAP directory, has an object class of `inetorgperson`, and did specify a valid password.

## Specifying a complete filter

A complete filter must contain a `@USERNAME@` token which will be substituted with the user name during the authentication attempt. The `rgw_ldap_dnattr` parameter is not used anymore in this case. For example, to limit valid users to a specific group, use the following filter:

```
1. "(&(uid=@USERNAME@)(memberOf=cn=ceph-users,ou=groups,dc=mycompany,dc=com))"
```

### Note

Using the `memberof` attribute in LDAP searches requires server side support from your specific LDAP server implementation.

## Generating an access token for LDAP authentication

The `radosgw-token` utility generates the access token based on the LDAP user name and password. It will output a base-64 encoded string which is the access token.

```
1. # export RGW_ACCESS_KEY_ID=<username>
2. # export RGW_SECRET_ACCESS_KEY=<password>
3. # radosgw-token --encode
```

### Important

The access token is a base-64 encoded JSON struct and contains the LDAP credentials as a clear text.

Alternatively, users can also generate the token manually by base-64-encoding this JSON snippet, if they do not have the `radosgw-token` tool installed.

```
1. {
2.   "RGW_TOKEN": {
3.     "version": 1,
4.     "type": "ldap",
```

```
5.     "id": "your_username",
6.     "key": "your_clear_text_password_here"
7.   }
8. }
```

## Using the access token

Use your favorite S3 client and specify the token as the access key in your client or environment variables.

```
1. # export AWS_ACCESS_KEY_ID=<base64-encoded token generated by radosgw-token>
  # export AWS_SECRET_ACCESS_KEY="" # define this with an empty string, otherwise tools might complain about
  2. missing env variables.
```

### Important

The access token is a base-64 encoded JSON struct and contains the LDAP credentials as a clear text. DO NOT share it unless you want to share your clear text password!

# Encryption

New in version Luminous.

The Ceph Object Gateway supports server-side encryption of uploaded objects, with 3 options for the management of encryption keys. Server-side encryption means that the data is sent over HTTP in its unencrypted form, and the Ceph Object Gateway stores that data in the Ceph Storage Cluster in encrypted form.

## Note

Requests for server-side encryption must be sent over a secure HTTPS connection to avoid sending secrets in plaintext. If a proxy is used for SSL termination, `rgw trust forwarded https` must be enabled before forwarded requests will be trusted as secure.

## Note

Server-side encryption keys must be 256-bit long and base64 encoded.

## Customer-Provided Keys

In this mode, the client passes an encryption key along with each request to read or write encrypted data. It is the client's responsibility to manage those keys and remember which key was used to encrypt each object.

This is implemented in S3 according to the [Amazon SSE-C](#) specification.

As all key management is handled by the client, no special configuration is needed to support this encryption mode.

## Key Management Service

This mode allows keys to be stored in a secure key management service and retrieved on demand by the Ceph Object Gateway to serve requests to encrypt or decrypt data.

This is implemented in S3 according to the [Amazon SSE-KMS](#) specification.

In principle, any key management service could be used here, but currently only integration with [Barbican](#) and [Vault](#) are implemented.

See [OpenStack Barbican Integration](#) and [HashiCorp Vault Integration](#).

## Automatic Encryption (for testing only)

A `rgw crypt default encryption key` can be set in ceph.conf to force the encryption of all objects that do not otherwise specify an encryption mode.

The configuration expects a base64-encoded 256 bit key. For example:

```
1. rgw crypt default encryption key = 4YSmvJtBv0aZ7geVgAsdpRnLBElWSWlMIGnRS8a9TSA=
```

### Important

This mode is for diagnostic purposes only! The ceph configuration file is not a secure method for storing encryption keys. Keys that are accidentally exposed in this way should be considered compromised.

# Bucket Policies

New in version Luminous.

The Ceph Object Gateway supports a subset of the Amazon S3 policy language applied to buckets.

## Creation and Removal

Bucket policies are managed through standard S3 operations rather than radosgw-admin.

For example, one may use s3cmd to set or delete a policy thus:

```
1. $ cat > examplepol
2. {
3.     "Version": "2012-10-17",
4.     "Statement": [
5.         "Effect": "Allow",
6.         "Principal": {"AWS": ["arn:aws:iam::usfolks:user/fred:subuser"]},
7.         "Action": "s3:PutObjectAcl",
8.         "Resource": [
9.             "arn:aws:s3:::happybucket/*"
10.        ]
11.    ]
12. }
13.
14. $ s3cmd setpolicy examplepol s3://happybucket
15. $ s3cmd delpolicy s3://happybucket
```

## Limitations

Currently, we support only the following actions:

- s3:AbortMultipartUpload
- s3:CreateBucket
- s3>DeleteBucketPolicy
- s3>DeleteBucket
- s3>DeleteBucketWebsite
- s3>DeleteObject
- s3>DeleteObjectVersion
- s3>DeleteReplicationConfiguration

- s3:GetAccelerateConfiguration
- s3:GetBucketAcl
- s3:GetBucketCORS
- s3:GetBucketLocation
- s3:GetBucketLogging
- s3:GetBucketNotification
- s3:GetBucketPolicy
- s3:GetBucketRequestPayment
- s3:GetBucketTagging
- s3:GetBucketVersioning
- s3:GetBucketWebsite
- s3:GetLifecycleConfiguration
- s3:GetObjectAcl
- s3:GetObject
- s3:GetObjectTorrent
- s3:GetObjectVersionAcl
- s3:GetObjectVersion
- s3:GetObjectVersionTorrent
- s3:GetReplicationConfiguration
- s3:IPAddress
- s3:NotIpAddress
- s3>ListAllMyBuckets
- s3>ListBucketMultipartUploads
- s3>ListBucket
- s3>ListBucketVersions
- s3>ListMultipartUploadParts
- s3PutAccelerateConfiguration

- s3:PutBucketAcl
- s3:PutBucketCORS
- s3:PutBucketLogging
- s3:PutBucketNotification
- s3:PutBucketPolicy
- s3:PutBucketRequestPayment
- s3:PutBucketTagging
- s3:PutBucketVersioning
- s3:PutBucketWebsite
- s3:PutLifecycleConfiguration
- s3:PutObjectAcl
- s3:PutObject
- s3:PutObjectVersionAcl
- s3:PutReplicationConfiguration
- s3:RestoreObject

We do not yet support setting policies on users, groups, or roles.

We use the RGW ‘tenant’ identifier in place of the Amazon twelve-digit account ID. In the future we may allow you to assign an account ID to a tenant, but for now if you want to use policies between AWS S3 and RGW S3 you will have to use the Amazon account ID as the tenant ID when creating users.

Under AWS, all tenants share a single namespace. RGW gives every tenant its own namespace of buckets. There may be an option to enable an AWS-like ‘flat’ bucket namespace in future versions. At present, to access a bucket belonging to another tenant, address it as “tenant:bucket” in the S3 request.

In AWS, a bucket policy can grant access to another account, and that account owner can then grant access to individual users with user permissions. Since we do not yet support user, role, and group permissions, account owners will currently need to grant access directly to individual users, and granting an entire account access to a bucket grants access to all users in that account.

Bucket policies do not yet support string interpolation.

For all requests, condition keys we support are: - aws:CurrentTime - aws:EpochTime - aws:PrincipalType - aws:Referer - aws:SecureTransport - aws:SourceIp - aws:UserAgent -

aws:username

We support certain s3 condition keys for bucket and object requests.

New in version Mimic.

## Bucket Related Operations

| Permission                               | Condition Keys                                                                                      | Comments |
|------------------------------------------|-----------------------------------------------------------------------------------------------------|----------|
| s3:createBucket                          | s3:x-amz-acl s3:x-amz-grant-<perm> where perm is one of read/write/read-acp write-acp/ full-control |          |
| s3>ListBucket &<br>s3>ListBucketVersions | s3:prefix                                                                                           |          |
|                                          | s3:delimiter                                                                                        |          |
|                                          | s3:max-keys                                                                                         |          |
| s3:PutBucketAcl                          | s3:x-amz-acl s3:x-amz-grant-<perm>                                                                  |          |

## Object Related Operations

| Permission                                | Condition Keys                                 | Comments                                                   |
|-------------------------------------------|------------------------------------------------|------------------------------------------------------------|
| s3:PutObject                              | s3:x-amz-acl & s3:x-amz-grant-<perm>           |                                                            |
|                                           | s3:x-amz-copy-source                           |                                                            |
|                                           | s3:x-amz-server-side-encryption                |                                                            |
|                                           | s3:x-amz-server-side-encryption-aws-kms-key-id |                                                            |
|                                           | s3:x-amz-metadata-directive                    | PUT & COPY to overwrite/preserve metadata in COPY requests |
| s3:PutObjectAcl<br>s3:PutObjectVersionAcl | s3:RequestObjectTag/<tag-key>                  |                                                            |
|                                           | s3:x-amz-acl & s3:x-amz-grant-<perm>           |                                                            |
|                                           | s3:ExistingObjectTag/<tag-                     |                                                            |

|                                                                                |                                                   |  |
|--------------------------------------------------------------------------------|---------------------------------------------------|--|
|                                                                                | <code>key&gt;</code>                              |  |
| <code>s3:PutObjectTagging &amp;<br/>s3:PutObjectVersionTagging</code>          | <code>s3:RequestObjectTag/&lt;tag-key&gt;</code>  |  |
|                                                                                | <code>s3:ExistingObjectTag/&lt;tag-key&gt;</code> |  |
| <code>s3:GetObject &amp;<br/>s3:GetObjectVersion</code>                        | <code>s3:ExistingObjectTag/&lt;tag-key&gt;</code> |  |
| <code>s3:GetObjectAcl &amp;<br/>s3:GetObjectVersionAcl</code>                  | <code>s3:ExistingObjectTag/&lt;tag-key&gt;</code> |  |
| <code>s3:GetObjectTagging &amp;<br/>s3:GetObjectVersionTagging</code>          | <code>s3:ExistingObjectTag/&lt;tag-key&gt;</code> |  |
| <code>s3&gt;DeleteObjectTagging &amp;<br/>s3:DeleteObjectVersionTagging</code> | <code>s3:ExistingObjectTag/&lt;tag-key&gt;</code> |  |

More may be supported soon as we integrate with the recently rewritten Authentication/Authorization subsystem.

## Swift

There is no way to set bucket policies under Swift, but bucket policies that have been set govern Swift as well as S3 operations.

Swift credentials are matched against Principals specified in a policy in a way specific to whatever backend is being used.

# RGW Dynamic Bucket Index Resharding

New in version Luminous.

A large bucket index can lead to performance problems. In order to address this problem we introduced bucket index sharding. Until Luminous, changing the number of bucket shards (resharding) needed to be done offline. Starting with Luminous we support online bucket resharding.

Each bucket index shard can handle its entries efficiently up until reaching a certain threshold number of entries. If this threshold is exceeded the system can suffer from performance issues. The dynamic resharding feature detects this situation and automatically increases the number of shards used by the bucket index, resulting in a reduction of the number of entries in each bucket index shard. This process is transparent to the user.

By default dynamic bucket index resharding can only increase the number of bucket index shards to 1999, although this upper-bound is a configuration parameter (see Configuration below). When possible, the process chooses a prime number of bucket index shards to spread the number of bucket index entries across the bucket index shards more evenly.

The detection process runs in a background process that periodically scans all the buckets. A bucket that requires resharding is added to the resharding queue and will be scheduled to be resharded later. The reshards thread runs in the background and execute the scheduled resharding tasks, one at a time.

## Multisite

Dynamic resharding is not supported in a multisite environment.

## Configuration

Enable/Disable dynamic bucket index resharding:

- `rgw_dynamic_resharding` : true/false, default: true

Configuration options that control the resharding process:

- `rgw_max_objs_per_shard` : maximum number of objects per bucket index shard before resharding is triggered, default: 100000 objects
- `rgw_max_dynamic_shards` : maximum number of shards that dynamic bucket index resharding can increase to, default: 1999
- `rgw_reshard_bucket_lock_duration` : duration, in seconds, of lock on bucket obj during

resharding, default: 360 seconds (i.e., 6 minutes)

- `rgw_reshard_thread_interval` : maximum time, in seconds, between rounds of resharding queue processing, default: 600 seconds (i.e., 10 minutes)
- `rgw_reshard_num_logs` : number of shards for the resharding queue, default: 16

## Admin commands

### Add a bucket to the resharding queue

```
1. # radosgw-admin reshard add --bucket <bucket_name> --num-shards <new number of shards>
```

### List resharding queue

```
1. # radosgw-admin reshard list
```

### Process tasks on the resharding queue

```
1. # radosgw-admin reshard process
```

### Bucket resharding status

```
1. # radosgw-admin reshard status --bucket <bucket_name>
```

The output is a json array of 3 objects (`reshard_status`, `new_bucket_instance_id`, `num_shards`) per shard.

For example, the output at different Dynamic Resharding stages is shown below:

1. Before resharding occurred:

```
1. [
2.   {
3.     "reshard_status": "not-resharding",
4.     "new_bucket_instance_id": "",
5.     "num_shards": -1
6.   }
7. ]
```

2. During resharding:

```
1. [
2.   {
```

```

3.     "reshard_status": "in-progress",
4.     "new_bucket_instance_id": "1179f470-2ebf-4630-8ec3-c9922da887fd.8652.1",
5.     "num_shards": 2
6.   },
7.   {
8.     "reshard_status": "in-progress",
9.     "new_bucket_instance_id": "1179f470-2ebf-4630-8ec3-c9922da887fd.8652.1",
10.    "num_shards": 2
11.  }
12. ]

```

3, After reshading completed:

```

1. [
2.   {
3.     "reshard_status": "not-resharding",
4.     "new_bucket_instance_id": "",
5.     "num_shards": -1
6.   },
7.   {
8.     "reshard_status": "not-resharding",
9.     "new_bucket_instance_id": "",
10.    "num_shards": -1
11.  }
12. ]

```

## Cancel pending bucket reshading

Note: Ongoing bucket reshading operations cannot be cancelled.

```
1. # radosgw-admin reshade cancel --bucket <bucket_name>
```

## Manual immediate bucket reshading

```
1. # radosgw-admin bucket reshade --bucket <bucket_name> --num-shards <new number of shards>
```

When choosing a number of shards, the administrator should keep a number of items in mind. Ideally the administrator is aiming for no more than 100000 entries per shard, now and through some future point in time.

Additionally, bucket index shards that are prime numbers tend to work better in evenly distributing bucket index entries across the shards. For example, 7001 bucket index shards is better than 7000 since the former is prime. A variety of web sites have lists of prime numbers; search for “list of prime numbers” with your favorite web search engine to locate some web sites.

## Troubleshooting

Clusters prior to Luminous 12.2.11 and Mimic 13.2.5 left behind stale bucket instance entries, which were not automatically cleaned up. The issue also affected LifeCycle policies, which were not applied to resharded buckets anymore. Both of these issues can be worked around using a couple of radosgw-admin commands.

## Stale instance management

List the stale instances in a cluster that are ready to be cleaned up.

```
1. # radosgw-admin reshard stale-instances list
```

Clean up the stale instances in a cluster. Note: cleanup of these instances should only be done on a single site cluster.

```
1. # radosgw-admin reshard stale-instances rm
```

## Lifecycle fixes

For clusters that had resharded instances, it is highly likely that the old lifecycle processes would have flagged and deleted lifecycle processing as the bucket instance changed during a reshard. While this is fixed for newer clusters (from Mimic 13.2.6 and Luminous 12.2.12), older buckets that had lifecycle policies and that have undergone resharding will have to be manually fixed.

The command to do so is:

```
1. # radosgw-admin lc reshard fix --bucket {bucketname}
```

As a convenience wrapper, if the `--bucket` argument is dropped then this command will try and fix lifecycle policies for all the buckets in the cluster.

## Object Expirer fixes

Objects subject to Swift object expiration on older clusters may have been dropped from the log pool and never deleted after the bucket was resharded. This would happen if their expiration time was before the cluster was upgraded, but if their expiration was after the upgrade the objects would be correctly handled. To manage these expire-stale objects, radosgw-admin provides two subcommands.

Listing:

```
1. # radosgw-admin objects expire-stale list --bucket {bucketname}
```

Displays a list of object names and expiration times in JSON format.

Deleting:

```
1. # radosgw-admin objects expire-stale rm --bucket {bucketname}
```

Initiates deletion of such objects, displaying a list of object names, expiration times, and deletion status in JSON format.

# RGW Support for Multifactor Authentication

New in version Mimic.

The S3 multifactor authentication (MFA) feature allows users to require the use of one-time password when removing objects on certain buckets. The buckets need to be configured with versioning and MFA enabled which can be done through the S3 api.

Time-based one time password tokens can be assigned to a user through radosgw-admin. Each token has a secret seed, and a serial id that is assigned to it. Tokens are added to the user, can be listed, removed, and can also be re-synchronized.

## Multisite

While the MFA IDs are set on the user's metadata, the actual MFA one time password configuration resides in the local zone's osds. Therefore, in a multi-site environment it is advisable to use different tokens for different zones.

## Terminology

- **TOTP** : Time-based One Time Password
- **token serial** : a string that represents the ID of a TOTP token
- **token seed** : the secret seed that is used to calculate the TOTP
- **totp seconds** : the time resolution that is being used for TOTP generation
- **totp window** : the number of TOTP tokens that are checked before and after the current token when validating token
- **totp pin** : the valid value of a TOTP token at a certain time

## Admin commands

### Create a new MFA TOTP token

```

1. # radosgw-admin mfa create --uid=<user-id> \
2.                               --totp-serial=<serial> \
3.                               --totp-seed=<seed> \
4.                               [ --totp-seed-type=<hex|base32> ] \
5.                               [ --totp-seconds=<num-seconds> ] \
6.                               [ --totp-window=<twindow> ]

```

## List MFA TOTP tokens

```
1. # radosgw-admin mfa list --uid=<user-id>
```

## Show MFA TOTP token

```
1. # radosgw-admin mfa get --uid=<user-id> --totp-serial=<serial>
```

## Delete MFA TOTP token

```
1. # radosgw-admin mfa remove --uid=<user-id> --totp-serial=<serial>
```

## Check MFA TOTP token

Test a TOTP token pin, needed for validating that TOTP functions correctly.

```
1. # radosgw-admin mfa check --uid=<user-id> --totp-serial=<serial> \
2.           --totp-pin=<pin>
```

## Re-sync MFA TOTP token

In order to re-sync the TOTP token (in case of time skew). This requires feeding two consecutive pins: the previous pin, and the current pin.

```
1. # radosgw-admin mfa resync --uid=<user-id> --totp-serial=<serial> \
2.           --totp-pin=<prev-pin> --totp-pin=<current-pin>
```

# Sync Modules

New in version Kraken.

The [Multi-Site](#) functionality of RGW introduced in Jewel allowed the ability to create multiple zones and mirror data and metadata between them. [Sync Modules](#) are built atop of the multisite framework that allows for forwarding data and metadata to a different external tier. A sync module allows for a set of actions to be performed whenever a change in data occurs (metadata ops like bucket or user creation etc. are also regarded as changes in data). As the rgw multisite changes are eventually consistent at remote sites, changes are propagated asynchronously. This would allow for unlocking use cases such as backing up the object storage to an external cloud cluster or a custom backup solution using tape drives, indexing metadata in ElasticSearch etc.

A sync module configuration is local to a zone. The sync module determines whether the zone exports data or can only consume data that was modified in another zone. As of luminous the supported sync plugins are [elasticsearch](#), [rgw](#), which is the default sync plugin that synchronises data between the zones and [log](#) which is a trivial sync plugin that logs the metadata operation that happens in the remote zones. The following docs are written with the example of a zone using [elasticsearch sync module](#), the process would be similar for configuring any sync plugin

- [ElasticSearch Sync Module](#)
- [Cloud Sync Module](#)
- [PubSub Module](#)
- [Archive Sync Module](#)

## Requirements and Assumptions

Let us assume a simple multisite configuration as described in the [Multi-Site](#) docs, of 2 zones [us-east](#) and [us-west](#), let's add a third zone [us-east-es](#) which is a zone that only processes metadata from the other sites. This zone can be in the same or a different ceph cluster as [us-east](#). This zone would only consume metadata from other zones and RGWs in this zone will not serve any end user requests directly.

## Configuring Sync Modules

Create the third zone similar to the [Multi-Site](#) docs, for example

```
1. # radosgw-admin zone create --rgw-zonegroup=us --rgw-zone=us-east-es \
2. --access-key={system-key} --secret={secret} --endpoints=http://rgw-es:80
```

A sync module can be configured for this zone via the following

```
# radosgw-admin zone modify --rgw-zone={zone-name} --tier-type={tier-type} --tier-config={set of key=value pairs}
```

For example in the `elasticsearch` sync module

```
1. # radosgw-admin zone modify --rgw-zone={zone-name} --tier-type=elasticsearch \
2.           --tier-config=endpoint=http://localhost:9200,num_shards=10,num_replicas=1
```

For the various supported tier-config options refer to the [elasticsearch sync module](#) docs

Finally update the period

```
1. # radosgw-admin period update --commit
```

Now start the radosgw in the zone

```
1. # systemctl start ceph-radosgw@rgw.`hostname -s` \
2. # systemctl enable ceph-radosgw@rgw.`hostname -s`
```

# ElasticSearch Sync Module

New in version Kraken.

This sync module writes the metadata from other zones to [ElasticSearch](#). As of luminous this is a json of data fields we currently store in ElasticSearch.

```

1. {
2.     "_index" : "rgw-gold-ee5863d6",
3.     "_type" : "object",
4.     "_id" : "34137443-8592-48d9-8ca7-160255d52ade.34137.1:object1:null",
5.     "_score" : 1.0,
6.     "_source" : {
7.         "bucket" : "testbucket123",
8.         "name" : "object1",
9.         "instance" : "null",
10.        "versioned_epoch" : 0,
11.        "owner" : {
12.            "id" : "user1",
13.            "display_name" : "user1"
14.        },
15.        "permissions" : [
16.            "user1"
17.        ],
18.        "meta" : {
19.            "size" : 712354,
20.            "mtime" : "2017-05-04T12:54:16.462Z",
21.            "etag" : "7ac66c0f148de9519b8bd264312c4d64"
22.        }
23.    }
24. }
```

## ElasticSearch tier type configurables

- `endpoint`

Specifies the Elasticsearch server endpoint to access

- `num_shards` (integer)

The number of shards that Elasticsearch will be configured with on data sync initialization. Note that this cannot be changed after init. Any change here requires rebuild of the Elasticsearch index and reinit of the data sync process.

- `num_replicas` (integer)

The number of the replicas that Elasticsearch will be configured with on data sync initialization.

- `explicit_custom_meta` (true | false)

Specifies whether all user custom metadata will be indexed, or whether user will need to configure (at the bucket level) what custom metadata entries should be indexed. This is false by default

- `index_buckets_list` (comma separated list of strings)

If empty, all buckets will be indexed. Otherwise, only buckets specified here will be indexed. It is possible to provide bucket prefixes (e.g., `foo*`), or bucket suffixes (e.g., `*bar`).

- `approved_owners_list` (comma separated list of strings)

If empty, buckets of all owners will be indexed (subject to other restrictions), otherwise, only buckets owned by specified owners will be indexed. Suffixes and prefixes can also be provided.

- `override_index_path` (string)

if not empty, this string will be used as the elasticsearch index path. Otherwise the index path will be determined and generated on sync initialization.

## End user metadata queries

---

New in version Luminous.

Since the ElasticSearch cluster now stores object metadata, it is important that the ElasticSearch endpoint is not exposed to the public and only accessible to the cluster administrators. For exposing metadata queries to the end user itself this poses a problem since we'd want the user to only query their metadata and not of any other users, this would require the ElasticSearch cluster to authenticate users in a way similar to RGW does which poses a problem.

As of Luminous RGW in the metadata master zone can now service end user requests. This allows for not exposing the elasticsearch endpoint in public and also solves the authentication and authorization problem since RGW itself can authenticate the end user requests. For this purpose RGW introduces a new query in the bucket APIs that can service elasticsearch requests. All these requests must be sent to the metadata master zone.

## Syntax

### Get an elasticsearch query

```
1. GET /{bucket}?query={query-expr}
```

request params:

- max-keys: max number of entries to return
- marker: pagination marker

```
expression := [()<arg> <op> <value> []][<and|or> ...]
```

op is one of the following: <, <=, ==, >=, >

For example

```
1. GET /?query=name==foo
```

Will return all the indexed keys that user has read permission to, and are named 'foo'.

The output will be a list of keys in XML that is similar to the S3 list buckets response.

## Configure custom metadata fields

Define which custom metadata entries should be indexed (under the specified bucket), and what are the types of these keys. If explicit custom metadata indexing is configured, this is needed so that rgw will index the specified custom metadata values. Otherwise it is needed in cases where the indexed metadata keys are of a type other than string.

```
1. POST /{bucket}?mdsearch
2. x-amz-meta-search: <key [<type>] [, ...]
```

Multiple metadata fields must be comma separated, a type can be forced for a field with a ;. The currently allowed types are string(default), integer and date

eg. if you want to index a custom object metadata x-amz-meta-year as int, x-amz-meta-date as type date and x-amz-meta-title as string, you'd do

```
1. POST /mybooks?mdsearch
2. x-amz-meta-search: x-amz-meta-year;int, x-amz-meta-release-date;date, x-amz-meta-title;string
```

## Delete custom metadata configuration

Delete custom metadata bucket configuration.

```
1. DELETE /<bucket>?mdsearch
```

## Get custom metadata configuration

Retrieve custom metadata bucket configuration.

```
1. GET /<bucket>?mdsearch
```

# Cloud Sync Module

New in version Mimic.

This module syncs zone data to a remote cloud service. The sync is unidirectional; data is not synced back from the remote zone. The goal of this module is to enable syncing data to multiple cloud providers. The currently supported cloud providers are those that are compatible with AWS (S3).

User credentials for the remote cloud object store service need to be configured. Since many cloud services impose limits on the number of buckets that each user can create, the mapping of source objects and buckets is configurable. It is possible to configure different targets to different buckets and bucket prefixes. Note that source ACLs will not be preserved. It is possible to map permissions of specific source users to specific destination users.

Due to API limitations there is no way to preserve original object modification time and ETag. The cloud sync module stores these as metadata attributes on the destination objects.

## Cloud Sync Tier Type Configuration

### Trivial Configuration:

```
1. {
2.   "connection": {
3.     "access_key": <access>,
4.     "secret": <secret>,
5.     "endpoint": <endpoint>,
6.     "host_style": <path | virtual>,
7.   },
8.   "acls": [ { "type": <id | email | uri>,
9.             "source_id": <source_id>,
10.            "dest_id": <dest_id> } ... ],
11.   "target_path": <target_path>,
12. }
```

### Non Trivial Configuration:

```
1. {
2.   "default": {
3.     "connection": {
4.       "access_key": <access>,
5.       "secret": <secret>,
6.       "endpoint": <endpoint>,
```

```

7.      "host_style" <path | virtual>,
8.    },
9.    "acls": [
10.    {
11.      "type" : <id | email | uri>,   # optional, default is id
12.      "source_id": <id>,
13.      "dest_id": <id>
14.    } ... ]
15.    "target_path": <path> # optional
16.  },
17.  "connections": [
18.    {
19.      "connection_id": <id>,
20.      "access_key": <access>,
21.      "secret": <secret>,
22.      "endpoint": <endpoint>,
23.      "host_style" <path | virtual>, # optional
24.    } ... ],
25.  "acl_profiles": [
26.    {
27.      "acls_id": <id>, # acl mappings
28.      "acls": [ {
29.          "type": <id | email | uri>,
30.          "source_id": <id>,
31.          "dest_id": <id>
32.        } ... ]
33.      }
34.    ],
35.    "profiles": [
36.      {
37.        "source_bucket": <source>,
38.        "connection_id": <connection_id>,
39.        "acls_id": <mappings_id>,
40.        "target_path": <dest>,       # optional
41.      } ... ],
42.  }

```

## Note

Trivial configuration can coincide with the non-trivial one.

- `connection` (container)

Represents a connection to the remote cloud service. Contains `conection_id`, `access_key`, `secret`, `endpoint`, and `host_style`.

- `access_key` (string)

The remote cloud access key that will be used for a specific connection.

- `secret` (string)

The secret key for the remote cloud service.

- `endpoint` (string)

URL of remote cloud service endpoint.

- `host_style` (path | virtual)

Type of host style to be used when accessing remote cloud endpoint (default: `path`).

- `acls` (array)

Contains a list of `acl_mappings`.

- `acl_mapping` (container)

Each `acl_mapping` structure contains `type`, `source_id`, and `dest_id`. These will define the ACL mutation that will be done on each object. An ACL mutation allows converting source user id to a destination id.

- `type` (id | email | uri)

ACL type: `id` defines user id, `email` defines user by email, and `uri` defines user by `uri` (group).

- `source_id` (string)

ID of user in the source zone.

- `dest_id` (string)

ID of user in the destination.

- `target_path` (string)

A string that defines how the target path is created. The target path specifies a prefix to which the source object name is appended. The target path configurable can include any of the following variables:

- `sid` : unique string that represents the sync instance ID
- `zonegroup` : the zonegroup name
- `zonegroup_id` : the zonegroup ID
- `zone` : the zone name
- `zone_id` : the zone id
- `bucket` : source bucket name
- `owner` : source bucket owner ID

For example: `target_path = rgwx-${zone}-${sid}/${owner}/${bucket}`

- `acl_profiles` (array)

An array of `acl_profile`.

- `acl_profile` (container)

Each profile contains `acls_id` (string) that represents the profile, and `acls` array that holds a list of `acl_mappings`.

- `profiles` (array)

A list of profiles. Each profile contains the following: - `source_bucket` : either a bucket name, or a bucket prefix (if ends with `*`) that defines the source bucket(s) for this profile - `target_path` : as defined above - `connection_id` : ID of the connection that will be used for this profile - `acls_id` : ID of ACLs profile that will be used for this profile

## S3 Specific Configurables:

Currently cloud sync will only work with backends that are compatible with AWS S3. There are a few configurables that can be used to tweak its behavior when accessing these cloud services:

```
1. {
2.   "multipart_sync_threshold": {object_size},
3.   "multipart_min_part_size": {part_size}
4. }
```

- `multipart_sync_threshold` (integer)

Objects this size or larger will be synced to the cloud using multipart upload.

- `multipart_min_part_size` (integer)

Minimum parts size to use when syncing objects using multipart upload.

## How to Configure

See [Multi-Site](#) for how to multisite config instructions. The cloud sync module requires a creation of a new zone. The zone tier type needs to be defined as `cloud` :

```
1. # radosgw-admin zone create --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --endpoints={http://fqdn}[,{http://fqdn}]
4.                               --tier-type=cloud
```

The tier configuration can be then done using the following command

```
1. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --tier-config={key}={val}[,{key}={val}]
```

The `key` in the configuration specifies the config variable that needs to be updated, and the `val` specifies its new value. Nested values can be accessed using period. For example:

```
1. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --tier-config=connection.access_key={key},connection.secret={secret}
```

Configuration array entries can be accessed by specifying the specific entry to be referenced enclosed in square brackets, and adding new array entry can be done by using []. Index value of -1 references the last entry in the array. At the moment it is not possible to create a new entry and reference it again at the same command. For example, creating a new profile for buckets starting with {prefix}:

```
1. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
   --rgw-zone={zone-name} \
   --tier-config=profiles[] .source_bucket={prefix} '*'
4.
5. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
   --rgw-zone={zone-name} \
   --tier-config=profiles[-1] .connection_id={conn_id}, profiles[-1] .acls_id={acls_id}
```

An entry can be removed by using `--tier-config-rm={key}` .

# PubSub Sync Module

New in version Nautilus.

## Contents

- PubSub Sync Module
  - PubSub Zone Configuration
    - PubSub Zone Configuration Parameters
    - Configuring Parameters via CLI
  - Topic and Subscription Management via CLI
  - PubSub Performance Stats
  - PubSub REST API
    - Topics
      - Create a Topic
      - Get Topic Information
      - Delete Topic
      - List Topics
    - S3-Compliant Notifications
    - Non S3-Compliant Notifications
      - Create a Notification
      - Delete Notification Information
      - List Notifications
    - Subscriptions
      - Create a Subscription
      - Get Subscription Information
      - Delete Subscription
    - Events
      - Pull Events

- [Ack Event](#)

This sync module provides a publish and subscribe mechanism for the object store modification events. Events are published into predefined topics. Topics can be subscribed to, and events can be pulled from them. Events need to be acked. Also, events will expire and disappear after a period of time.

A push notification mechanism exists too, currently supporting HTTP, AMQP0.9.1 and Kafka endpoints. In this case, the events are pushed to an endpoint on top of storing them in Ceph. If events should only be pushed to an endpoint and do not need to be stored in Ceph, the [Bucket Notification](#) mechanism should be used instead of pubsub sync module.

A user can create different topics. A topic entity is defined by its name and is per tenant. A user can only associate its topics (via notification configuration) with buckets it owns.

In order to publish events for specific bucket a notification entity needs to be created. A notification can be created on a subset of event types, or for all event types (default). There can be multiple notifications for any specific topic, and the same topic could be used for multiple notifications.

A subscription to a topic can also be defined. There can be multiple subscriptions for any specific topic.

REST API has been defined to provide configuration and control interfaces for the pubsub mechanisms. This API has two flavors, one is S3-compatible and one is not. The two flavors can be used together, although it is recommended to use the S3-compatible one. The S3-compatible API is similar to the one used in the bucket notification mechanism.

Events are stored as RGW objects in a special bucket, under a special user (pubsub control user). Events cannot be accessed directly, but need to be pulled and acked using the new REST API.

- [S3 Bucket Notification Compatibility](#)

## PubSub Zone Configuration

The pubsub sync module requires the creation of a new zone in a [Multi-Site](#) environment... First, a master zone must exist (see: [Configuring a Master Zone](#)), then a secondary zone should be created (see [Configure Secondary Zones](#)). In the creation of the secondary zone, its tier type must be set to `pubsub` :

```
1. # radosgw-admin zone create --rgw-zonegroup={zone-group-name} \
2.           --rgw-zone={zone-name} \
3.           --endpoints={http://fqdn}[,{http://fqdn}] \
4.           --sync-from-all=0 \
5.           --sync-from={master-zone-name} \
```

```
6.          --tier-type=pubsub
```

## PubSub Zone Configuration Parameters

```
1. {
2.     "tenant": <tenant>,           # default: <empty>
3.     "uid": <uid>,                 # default: "pubsub"
4.     "data_bucket_prefix": <prefix> # default: "pubsub-"
5.     "data_oid_prefix": <prefix>   #
6.     "events_retention_days": <days> # default: 7
7. }
```

- `tenant` (string)

The tenant of the pubsub control user.

- `uid` (string)

The uid of the pubsub control user.

- `data_bucket_prefix` (string)

The prefix of the bucket name that will be created to store events for specific topic.

- `data_oid_prefix` (string)

The oid prefix for the stored events.

- `events_retention_days` (integer)

How many days to keep events that weren't acked.

## Configuring Parameters via CLI

The tier configuration could be set using the following command:

```
1. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --tier-config={key}={val}[,{key}={val}]
```

Where the `key` in the configuration specifies the configuration variable that needs to be updated (from the list above), and the `val` specifies its new value. For example, setting the pubsub control user `uid` to `user_ps` :

```
1. # radosgw-admin zone modify --rgw-zonegroup={zone-group-name} \
2.                               --rgw-zone={zone-name} \
3.                               --tier-config=uid=pubsub
```

A configuration field can be removed by using `--tier-config-rm={key}`.

# Topic and Subscription Management via CLI

Configuration of all topics, associated with a tenant, could be fetched using the following command:

```
1. # radosgw-admin topic list [--tenant={tenant}]
```

Configuration of a specific topic could be fetched using:

```
1. # radosgw-admin topic get --topic={topic-name} [--tenant={tenant}]
```

And removed using:

```
1. # radosgw-admin topic rm --topic={topic-name} [--tenant={tenant}]
```

Configuration of a subscription could be fetched using:

```
1. # radosgw-admin subscription get --subscription={topic-name} [--tenant={tenant}]
```

And removed using:

```
1. # radosgw-admin subscription rm --subscription={topic-name} [--tenant={tenant}]
```

To fetch all of the events stored in a subscription, use:

```
1. # radosgw-admin subscription pull --subscription={topic-name} [--marker={last-marker}] [--tenant={tenant}]
```

To ack (and remove) an event from a subscription, use:

```
1. # radosgw-admin subscription ack --subscription={topic-name} --event-id={event-id} [--tenant={tenant}]
```

## PubSub Performance Stats

Same counters are shared between the pubsub sync module and the notification mechanism.

- `pubsub_event_triggered` : running counter of events with at least one topic associated with them
- `pubsub_event_lost` : running counter of events that had topics and subscriptions associated with them but that were not stored or pushed to any of the subscriptions
- `pubsub_store_ok` : running counter, for all subscriptions, of stored events

- `pubsub_store_fail` : running counter, for all subscriptions, of events failed to be stored
- `pubsub_push_ok` : running counter, for all subscriptions, of events successfully pushed to their endpoint
- `pubsub_push_fail` : running counter, for all subscriptions, of events failed to be pushed to their endpoint
- `pubsub_push_pending` : gauge value of events pushed to an endpoint but not acked or nacked yet

#### Note

`pubsub_event_triggered` and `pubsub_event_lost` are incremented per event, while:  
`pubsub_store_ok`, `pubsub_store_fail`, `pubsub_push_ok`, `pubsub_push_fail`, are incremented per store/push action on each subscriptions.

## PubSub REST API

#### Tip

PubSub REST calls, and only them, should be sent to an RGW which belong to a PubSub zone

## Topics

### Create a Topic

This will create a new topic. Topic creation is needed both for both flavors of the API. Optionally the topic could be provided with push endpoint parameters that would be used later when an S3-compatible notification is created. Upon successful request, the response will include the topic ARN that could be later used to reference this topic in an S3-compatible notification request. To update a topic, use the same command used for topic creation, with the topic name of an existing topic and different endpoint values.

#### Tip

Any S3-compatible notification already associated with the topic needs to be re-created for the topic update to take effect

```
PUT /topics/<topic-name>[?OpaqueData=<opaque data>][&push-endpoint=<endpoint>[&amqp-exchange=<exchange>]
[&amqp-ack-level=none|broker|routable][&verify-ssl=true|false][&kafka-ack-level=none|broker][&use-
1. ssl=true|false][&ca-location=<file path>]]
```

Request parameters:

- `push-endpoint`: URI of an endpoint to send push notification to

- **OpaqueData:** opaque data is set in the topic configuration and added to all notifications triggered by the topic

The endpoint URI may include parameters depending with the type of endpoint:

- **HTTP endpoint**

- URI: `http[s]://<fqdn>[:<port>]`
- port defaults to: 80/443 for HTTP/S accordingly
- verify-ssl: indicate whether the server certificate is validated by the client or not ("true" by default)

- **AMQP0.9.1 endpoint**

- URI: `amqp://[<user>:<password>@]<fqdn>[:<port>][/<vhost>]`
- user/password defaults to: guest/guest
- user/password may only be provided over HTTPS. Topic creation request will be rejected if not
- port defaults to: 5672
- vhost defaults to: "/"
- amqp-exchange: the exchanges must exist and be able to route messages based on topics (mandatory parameter for AMQP0.9.1). Different topics pointing to the same endpoint must use the same exchange
- amqp-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Three ack methods exist:
  - "none": message is considered "delivered" if sent to broker
  - "broker": message is considered "delivered" if acked by broker (default)
  - "routable": message is considered "delivered" if broker can route to a consumer

## Tip

The topic-name (see [Create a Topic](#)) is used for the AMQP topic ("routing key" for a topic exchange)

- **Kafka endpoint**

- URI: `kafka://[<user>:<password>@]<fqdn>[:<port>]`
- if `use-ssl` is set to "true", secure connection will be used for connecting with the broker ("false" by default)
- if `ca-location` is provided, and secure connection is used, the specified CA will be used, instead of the default one, to authenticate the broker

- user/password may only be provided over HTTPS. Topic creation request will be rejected if not
- user/password may only be provided together with `use-ssl`, connection to the broker would fail if not
- port defaults to: 9092
- kafka-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Two ack methods exist:
  - "none": message is considered "delivered" if sent to broker
  - "broker": message is considered "delivered" if acked by broker (default)

The topic ARN in the response will have the following format:

1. `arn:aws:sns:<zone-group>:<tenant>:<topic>`

## Get Topic Information

Returns information about specific topic. This includes subscriptions to that topic, and push-endpoint information, if provided.

1. `GET /topics/<topic-name>`

Response will have the following format (JSON):

```

1. {
2.   "topic": {
3.     "user": "",
4.     "name": "",
5.     "dest": {
6.       "bucket_name": "",
7.       "oid_prefix": "",
8.       "push_endpoint": "",
9.       "push_endpoint_args": "",
10.      "push_endpoint_topic": "",
11.      "stored_secret": "",
12.      "persistent": ""
13.    },
14.    "arn": ""
15.    "opaqueData": ""
16.  },
17.  "subs": []
18. }
```

- `topic.user`: name of the user that created the topic
- `name`: name of the topic

- dest.bucket\_name: not used
- dest.oid\_prefix: not used
- dest.push\_endpoint: in case of S3-compliant notifications, this value will be used as the push-endpoint URL
- if push-endpoint URL contain user/password information, request must be made over HTTPS. Topic get request will be rejected if not
- dest.push\_endpoint\_args: in case of S3-compliant notifications, this value will be used as the push-endpoint args
- dest.push\_endpoint\_topic: in case of S3-compliant notifications, this value will hold the topic name as sent to the endpoint (may be different than the internal topic name)
- topic.arn: topic ARN
- subs: list of subscriptions associated with this topic

## Delete Topic

```
1. DELETE /topics/<topic-name>
```

Delete the specified topic.

## List Topics

List all topics associated with a tenant.

```
1. GET /topics
```

- if push-endpoint URL contain user/password information, in any of the topic, request must be made over HTTPS. Topic list request will be rejected if not

## S3-Compliant Notifications

Detailed under: [Bucket Operations](#).

### Note

- Notification creation will also create a subscription for pushing/pulling events
- The generated subscription's name will have the same as the notification Id, and could be used later to fetch and ack events with the subscription API.
- Notification deletion will deletes all generated subscriptions

- In case that bucket deletion implicitly deletes the notification, the associated subscription will not be deleted automatically (any events of the deleted bucket could still be accessed), and will have to be deleted explicitly with the subscription deletion API
- Filtering based on metadata (which is an extension to S3) is not supported, and such rules will be ignored
- Filtering based on tags (which is an extension to S3) is not supported, and such rules will be ignored

## Non S3-Compliant Notifications

### Create a Notification

This will create a publisher for a specific bucket into a topic.

```
1. PUT /notifications/bucket/<bucket>?topic=<topic-name>[&events=<event>[, <event>]]
```

Request parameters:

- topic-name: name of topic
- event: event type (string), one of: `OBJECT_CREATE` , `OBJECT_DELETE` , `DELETE_MARKER_CREATE`

### Delete Notification Information

Delete publisher from a specific bucket into a specific topic.

```
1. DELETE /notifications/bucket/<bucket>?topic=<topic-name>
```

Request parameters:

- topic-name: name of topic

#### Note

When the bucket is deleted, any notification defined on it is also deleted

### List Notifications

List all topics with associated events defined on a bucket.

```
1. GET /notifications/bucket/<bucket>
```

Response will have the following format (JSON):

```
1. {"topics": [
```

```

2.     {
3.         "topic": {
4.             "user": "",
5.             "name": "",
6.             "dest": {
7.                 "bucket_name": "",
8.                 "oid_prefix": "",
9.                 "push_endpoint": "",
10.                "push_endpoint_args": "",
11.                "push_endpoint_topic": ""
12.            }
13.            "arn": ""
14.        },
15.        "events": []
16.    }
17. }

```

## Subscriptions

### Create a Subscription

Creates a new subscription.

```

PUT /subscriptions/<sub-name>?topic=<topic-name>[?push-endpoint=<endpoint>[&amqp-exchange=<exchange>][&amqp-
ack-level=none|broker|routable][&verify-ssl=true|false][&kafka-ack-level=none|broker][&ca-location=<file
1. path>]]

```

Request parameters:

- topic-name: name of topic
- push-endpoint: URI of endpoint to send push notification to

The endpoint URI may include parameters depending with the type of endpoint:

- HTTP endpoint

- URI: `http[s]://<fqdn>[:<port>]`
- port defaults to: 80/443 for HTTP/S accordingly
- verify-ssl: indicate whether the server certificate is validated by the client or not ("true" by default)

- AMQP0.9.1 endpoint

- URI: `amqp://[<user>:<password>@]<fqdn>[:<port>][/<vhost>]`
- user/password defaults to : guest/guest
- port defaults to: 5672

- vhost defaults to: “/”
- amqp-exchange: the exchanges must exist and be able to route messages based on topics (mandatory parameter for AMQP0.9.1)
- amqp-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Three ack methods exist:
  - “none”: message is considered “delivered” if sent to broker
  - “broker”: message is considered “delivered” if acked by broker (default)
  - “routable”: message is considered “delivered” if broker can route to a consumer

- Kafka endpoint

- URI: `kafka://[<user>:<password>@]<fqdn>[:<port>]`
- if `ca-location` is provided, secure connection will be used for connection with the broker
- user/password may only be provided over HTTPS. Topic creation request will be rejected if not
- user/password may only be provided together with `ca-location`. Topic creation request will be rejected if not
- port defaults to: 9092
- kafka-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Two ack methods exist:

- “none”: message is considered “delivered” if sent to broker
- “broker”: message is considered “delivered” if acked by broker (default)

## Get Subscription Information

Returns information about specific subscription.

1. GET `/subscriptions/<sub-name>`

Response will have the following format (JSON):

```

1. {
2.   "user":"",
3.   "name":"",
4.   "topic":"",
5.   "dest":{
6.     "bucket_name":"",
7.     "oid_prefix":"",
8.     "push_endpoint":",

```

```

9.      "push_endpoint_args":"",
10.     "push_endpoint_topic":"",
11.   }
12.   "s3_id":"",
13. }
```

- user: name of the user that created the subscription
- name: name of the subscription
- topic: name of the topic the subscription is associated with
- dest.bucket\_name: name of the bucket storing the events
- dest.oid\_prefix: oid prefix for the events stored in the bucket
- dest.push\_endpoint: in case of S3-compliant notifications, this value will be used as the push-endpoint URL
- if push-endpoint URL contain user/password information, request must be made over HTTPS. Topic get request will be rejected if not
- dest.push\_endpoint\_args: in case of S3-compliant notifications, this value will be used as the push-endpoint args
- dest.push\_endpoint\_topic: in case of S3-compliant notifications, this value will hold the topic name as sent to the endpoint (may be different than the internal topic name)
- s3\_id: in case of S3-compliant notifications, this will hold the notification name that created the subscription

## Delete Subscription

Removes a subscription.

```
1. DELETE /subscriptions/<sub-name>
```

## Events

### Pull Events

Pull events sent to a specific subscription.

```
1. GET /subscriptions/<sub-name>?events[&max-entries=<max-entries>][&marker=<marker>]
```

Request parameters:

- marker: pagination marker for list of events, if not specified will start from

the oldest

- max-entries: max number of events to return

The response will hold information on the current marker and whether there are more events not fetched:

```
1. {"next_marker":"","is_truncated": "",...}
```

The actual content of the response is depended with how the subscription was created. In case that the subscription was created via an S3-compatible notification, the events will have an S3-compatible record format (JSON):

```
1. {"Records": [
2.   {
3.     "eventVersion": "2.1",
4.     "eventSource": "aws:s3",
5.     "awsRegion": "",
6.     "eventTime": "",
7.     "eventName": "",
8.     "userIdentity": {
9.       "principalId": ""
10.    },
11.    "requestParameters": {
12.      "sourceIPAddress": ""
13.    },
14.    "responseElements": {
15.      "x-amz-request-id": "",
16.      "x-amz-id-2": ""
17.    },
18.    "s3": {
19.      "s3SchemaVersion": "1.0",
20.      "configurationId": "",
21.      "bucket": {
22.        "name": "",
23.        "ownerIdentity": {
24.          "principalId": ""
25.        },
26.        "arn": "",
27.        "id": ""
28.      },
29.      "object": {
30.        "key": "",
31.        "size": "0",
32.        "eTag": "",
33.        "versionId": "",
34.        "sequencer": "",
35.        "metadata": [],
36.        "tags": []
37.      }
38.    },
39.    "eventId": ""
40.  }
41. }
```

```

40.      "opaqueData":"",
41.    }
42. ]}

```

- awsRegion: zonegroup
- eventTime: timestamp indicating when the event was triggered
- eventName: either `s3:ObjectCreated:`, or `s3:ObjectRemoved:`
- userIdentity: not supported
- requestParameters: not supported
- responseElements: not supported
- s3.configurationId: notification ID that created the subscription for the event
- s3.bucket.name: name of the bucket
- s3.bucket.ownerIdentity.principalId: owner of the bucket
- s3.bucket.arn: ARN of the bucket
- s3.bucket.id: Id of the bucket (an extension to the S3 notification API)
- s3.object.key: object key
- s3.object.size: not supported
- s3.object.eTag: object etag
- s3.object.version: object version in case of versioned bucket
- s3.object.sequencer: monotonically increasing identifier of the change per object (hexadecimal format)
- s3.object.metadata: not supported (an extension to the S3 notification API)
- s3.object.tags: not supported (an extension to the S3 notification API)
- s3.eventId: unique ID of the event, that could be used for acking (an extension to the S3 notification API)
- s3.opaqueData: opaque data is set in the topic configuration and added to all notifications triggered by the topic (an extension to the S3 notification API)

In case that the subscription was not created via a non S3-compatible notification, the events will have the following event format (JSON):

```

1. {"events": [
2.   {
3.     "id": ""

```

```

4.      "event":"",
5.      "timestamp":"",
6.      "info":{
7.          "attrs":{
8.              "mtime": ""
9.          },
10.         "bucket":{
11.             "bucket_id":"",
12.             "name":"",
13.             "tenant": ""
14.         },
15.         "key":{
16.             "instance":"",
17.             "name": ""
18.         }
19.     }
20.   }
21. ]}

```

- id: unique ID of the event, that could be used for acking
- event: one of: `OBJECT_CREATE` , `OBJECT_DELETE` , `DELETE_MARKER_CREATE`
- timestamp: timestamp indicating when the event was sent
- info.attrs.mtime: timestamp indicating when the event was triggered
- info.bucket.bucket\_id: id of the bucket
- info.bucket.name: name of the bucket
- info.bucket.tenant: tenant the bucket belongs to
- info.key.instance: object version in case of versioned bucket
- info.key.name: object key

## Ack Event

Ack event so that it can be removed from the subscription history.

```
1. POST /subscriptions/<sub-name>?ack&event-id=<event-id>
```

Request parameters:

- event-id: id of event to be acked

# Archive Sync Module

New in version Nautilus.

This sync module leverages the versioning feature of the S3 objects in RGW to have an archive zone that captures the different versions of the S3 objects as they occur over time in the other zones.

An archive zone allows to have a history of versions of S3 objects that can only be eliminated through the gateways associated with the archive zone.

This functionality is useful to have a configuration where several non-versioned zones replicate their data and metadata through their zone gateways (mirror configuration) providing high availability to the end users, while the archive zone captures all the data updates and metadata for consolidate them as versions of S3 objects.

Including an archive zone in a multizone configuration allows you to have the flexibility of an S3 object history in one only zone while saving the space that the replicas of the versioned S3 objects would consume in the rest of the zones.

## Archive Sync Tier Type Configuration

### How to Configure

See [Multisite Configuration](#) for how to multisite config instructions. The archive sync module requires a creation of a new zone. The zone tier type needs to be defined as

```
archive :
```

```
1. # radosgw-admin zone create --rgw-zonegroup={zone-group-name} \
2.           --rgw-zone={zone-name} \
3.           --endpoints={http://fqdn}[,{http://fqdn}]
4.           --tier-type=archive
```

# Bucket Notifications

New in version Nautilus.

## Contents

- [Bucket Notifications](#)
  - [Notification Reliability](#)
  - [Topic Management via CLI](#)
  - [Notification Performance Stats](#)
  - [Bucket Notification REST API](#)
    - [Topics](#)
      - [Create a Topic](#)
      - [Get Topic Attributes](#)
      - [Get Topic Information](#)
      - [Delete Topic](#)
      - [List Topics](#)
    - [Notifications](#)
    - [Events](#)

Bucket notifications provide a mechanism for sending information out of the radosgw when certain events are happening on the bucket. Currently, notifications could be sent to: HTTP, AMQP0.9.1 and Kafka endpoints.

Note, that if the events should be stored in Ceph, in addition, or instead of being pushed to an endpoint, the [PubSub Module](#) should be used instead of the bucket notification mechanism.

A user can create different topics. A topic entity is defined by its name and is per tenant. A user can only associate its topics (via notification configuration) with buckets it owns.

In order to send notifications for events for a specific bucket, a notification entity needs to be created. A notification can be created on a subset of event types, or for all event types (default). The notification may also filter out events based on prefix/suffix and/or regular expression matching of the keys. As well as, on the metadata attributes attached to the object, or the object tags. There can be multiple notifications for any specific topic, and the same topic could be used for multiple

notifications.

REST API has been defined to provide configuration and control interfaces for the bucket notification mechanism. This API is similar to the one defined as the S3-compatible API of the pubsub sync module.

- [S3 Bucket Notification Compatibility](#)

## Notification Reliability

Notifications may be sent synchronously, as part of the operation that triggered them. In this mode, the operation is acked only after the notification is sent to the topic's configured endpoint, which means that the round trip time of the notification is added to the latency of the operation itself.

### Note

The original triggering operation will still be considered as successful even if the notification fail with an error, cannot be delivered or times out

Notifications may also be sent asynchronously. They will be committed into persistent storage and then asynchronously sent to the topic's configured endpoint. In this case, the only latency added to the original operation is of committing the notification to persistent storage.

### Note

If the notification fail with an error, cannot be delivered or times out, it will be retried until successfully acked

### Tip

To minimize the added latency in case of asynchronous notifications, it is recommended to place the "log" pool on fast media

## Topic Management via CLI

Configuration of all topics, associated with a tenant, could be fetched using the following command:

```
1. # radosgw-admin topic list [--tenant={tenant}]
```

Configuration of a specific topic could be fetched using:

```
1. # radosgw-admin topic get --topic={topic-name} [--tenant={tenant}]
```

And removed using:

```
1. # radosgw-admin topic rm --topic={topic-name} [--tenant={tenant}]
```

## Notification Performance Stats

The same counters are shared between the pubsub sync module and the bucket notification mechanism.

- `pubsub_event_triggered` : running counter of events with at least one topic associated with them
- `pubsub_event_lost` : running counter of events that had topics associated with them but that were not pushed to any of the endpoints
- `pubsub_push_ok` : running counter, for all notifications, of events successfully pushed to their endpoint
- `pubsub_push_fail` : running counter, for all notifications, of events failed to be pushed to their endpoint
- `pubsub_push_pending` : gauge value of events pushed to an endpoint but not acked or nacked yet

### Note

`pubsub_event_triggered` and `pubsub_event_lost` are incremented per event, while:  
`pubsub_push_ok` , `pubsub_push_fail` , are incremented per push action on each notification.

## Bucket Notification REST API

### Topics

#### Note

In all topic actions, the parameters are URL encoded, and sent in the message body using `application/x-www-form-urlencoded` content type

### Create a Topic

This will create a new topic. The topic should be provided with push endpoint parameters that would be used later when a notification is created. Upon a successful request, the response will include the topic ARN that could be later used to reference this topic in the notification request. To update a topic, use the same command used for topic creation, with the topic name of an existing topic and different endpoint values.

#### Tip

Any notification already associated with the topic needs to be re-created for the topic update to take effect

```

1. POST
2.
3. Action=CreateTopic
4. &Name=<topic-name>
5. [&Attributes.entry.1.key=amqp-exchange&Attributes.entry.1.value=<exchange>]
6. [&Attributes.entry.2.key=amqp-ack-level&Attributes.entry.2.value=none|broker|routable]
7. [&Attributes.entry.3.key=verify-ssl&Attributes.entry.3.value=true|false]
8. [&Attributes.entry.4.key=kafka-ack-level&Attributes.entry.4.value=none|broker]
9. [&Attributes.entry.5.key=use-ssl&Attributes.entry.5.value=true|false]
10. [&Attributes.entry.6.key=ca-location&Attributes.entry.6.value=<file path>]
11. [&Attributes.entry.7.key=OpaqueData&Attributes.entry.7.value=<opaque data>]
12. [&Attributes.entry.8.key=push-endpoint&Attributes.entry.8.value=<endpoint>]
13. [&Attributes.entry.9.key=persistent&Attributes.entry.9.value=true|false]
```

#### Request parameters:

- push-endpoint: URI of an endpoint to send push notification to
- OpaqueData: opaque data is set in the topic configuration and added to all notifications triggered by the topic
- persistent: indication whether notifications to this endpoint are persistent (=asynchronous) or not ("false" by default)
- HTTP endpoint

- URI: `http[s]://<fqdn>[:<port>]`
- port defaults to: 80/443 for HTTP/S accordingly
- verify-ssl: indicate whether the server certificate is validated by the client or not ("true" by default)

- AMQP0.9.1 endpoint

- URI: `amqp://[<user>:<password>@]<fqdn>[:<port>][/<vhost>]`
- user/password defaults to: guest/guest
- user/password may only be provided over HTTPS. If not, topic creation request will be rejected.
- port defaults to: 5672
- vhost defaults to: "/"
- amqp-exchange: the exchanges must exist and be able to route messages based on topics (mandatory parameter for AMQP0.9.1). Different topics pointing to the same endpoint must use the same exchange
- amqp-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Three ack methods exist:

- “none”: message is considered “delivered” if sent to broker
- “broker”: message is considered “delivered” if acked by broker (default)
- “routable”: message is considered “delivered” if broker can route to a consumer

**Tip**

The topic-name (see [Create a Topic](#)) is used for the AMQP topic (“routing key” for a topic exchange)

- Kafka endpoint

- URI: `kafka://[<user>:<password>@]<fqdn>[:<port>]`
- if `use-ssl` is set to “true”, secure connection will be used for connecting with the broker (“false” by default)
- if `ca-location` is provided, and secure connection is used, the specified CA will be used, instead of the default one, to authenticate the broker
- user/password may only be provided over HTTPS. If not, topic creation request will be rejected.
- user/password may only be provided together with `use-ssl`, if not, the connection to the broker would fail.
- port defaults to: 9092
- kafka-ack-level: no end2end acking is required, as messages may persist in the broker before delivered into their final destination. Two ack methods exist:
  - “none”: message is considered “delivered” if sent to broker
  - “broker”: message is considered “delivered” if acked by broker (default)

**Note**

- The key/value of a specific parameter does not have to reside in the same line, or in any specific order, but must use the same index
- Attribute indexing does not need to be sequential or start from any specific value
- [AWS Create Topic](#) has a detailed explanation of the endpoint attributes format. However, in our case different keys and values are used

The response will have the following format:

```
1. <CreateTopicResponse xmlns="https://sns.amazonaws.com/doc/2010-03-31/">
2.   <CreateTopicResult>
3.     <TopicArn></TopicArn>
```

```

4.      </CreateTopicResult>
5.      <ResponseMetadata>
6.          <RequestId></RequestId>
7.      </ResponseMetadata>
8.  </CreateTopicResponse>
```

The topic ARN in the response will have the following format:

1. arn:aws:sns:<zone-group>:<tenant>:<topic>

## Get Topic Attributes

Returns information about a specific topic. This includes push-endpoint information, if provided.

1. POST
- 2.
3. Action=GetTopicAttributes
4. &TopicArn=<topic-arn>

Response will have the following format:

```

1.  <GetTopicAttributesResponse>
2.      <GetTopicAttributesResult>
3.          <Attributes>
4.              <entry>
5.                  <key>User</key>
6.                  <value></value>
7.              </entry>
8.              <entry>
9.                  <key>Name</key>
10.                 <value></value>
11.             </entry>
12.             <entry>
13.                 <key>EndPoint</key>
14.                 <value></value>
15.             </entry>
16.             <entry>
17.                 <key>TopicArn</key>
18.                 <value></value>
19.             </entry>
20.             <entry>
21.                 <key>OpaqueData</key>
22.                 <value></value>
23.             </entry>
24.         </Attributes>
25.     </GetTopicAttributesResult>
26.     <ResponseMetadata>
27.         <RequestId></RequestId>
28.     </ResponseMetadata>
```

```
29. </GetTopicAttributesResponse>
```

- User: name of the user that created the topic
- Name: name of the topic
- EndPoint: JSON formatted endpoint parameters, including:
  - EndpointAddress: the push-endpoint URL
  - EndpointArgs: the push-endpoint args
  - EndpointTopic: the topic name that should be sent to the endpoint (may be different than the above topic name)
  - HasStoredSecret: “true” if endpoint URL contain user/password information. In this case request must be made over HTTPS. If not, topic get request will be rejected
  - Persistent: “true” is topic is persistent
- TopicArn: topic ARN
- OpaqueData: the opaque data set on the topic

## Get Topic Information

Returns information about specific topic. This includes push-endpoint information, if provided. Note that this API is now deprecated in favor of the AWS compliant GetTopicAttributes API.

```
1. POST
2.
3. Action=GetTopic
4. &TopicArn=<topic-arn>
```

Response will have the following format:

```
1. <GetTopicResponse>
2.   <GetTopicResult>
3.     <Topic>
4.       <User></User>
5.       <Name></Name>
6.       <EndPoint>
7.         <EndpointAddress></EndpointAddress>
8.         <EndpointArgs></EndpointArgs>
9.         <EndpointTopic></EndpointTopic>
10.        <HasStoredSecret></HasStoredSecret>
11.        <Persistent></Persistent>
12.      </EndPoint>
13.    <TopicArn></TopicArn>
```

```

14.      <OpaqueData></OpaqueData>
15.      </Topic>
16.    </GetTopicResult>
17.    <ResponseMetadata>
18.      <RequestId></RequestId>
19.    </ResponseMetadata>
20.  </GetTopicResponse>

```

- User: name of the user that created the topic
- Name: name of the topic
- EndpointAddress: the push-endpoint URL
- EndpointArgs: the push-endpoint args
- EndpointTopic: the topic name that should be sent to the endpoint (may be different than the above topic name)
- HasStoredSecret: “true” if endpoint URL contain user/password information. In this case request must be made over HTTPS. If not, topic get request will be rejected
- Persistent: “true” is topic is persistent
- TopicArn: topic ARN
- OpaqueData: the opaque data set on the topic

## Delete Topic

```

1. POST
2.
3. Action=DeleteTopic
4. &TopicArn=<topic-arn>

```

Delete the specified topic. Note that deleting a deleted topic should result with no-op and not a failure.

The response will have the following format:

```

1. <DeleteTopicResponse xmlns="https://sns.amazonaws.com/doc/2010-03-31/">
2.   <ResponseMetadata>
3.     <RequestId></RequestId>
4.   </ResponseMetadata>
5. </DeleteTopicResponse>

```

## List Topics

List all topics associated with a tenant.

1. POST
- 2.
3. Action=ListTopics

Response will have the following format:

```

1. <ListTopicdResponse xmlns="https://sns.amazonaws.com/doc/2010-03-31/">
2.   <ListTopicsRersult>
3.     <Topics>
4.       <member>
5.         <User></User>
6.         <Name></Name>
7.         <EndPoint>
8.           <EndpointAddress></EndpointAddress>
9.           <EndpointArgs></EndpointArgs>
10.          <EndpointTopic></EndpointTopic>
11.        </EndPoint>
12.        <TopicArn></TopicArn>
13.        <OpaqueData></OpaqueData>
14.      </member>
15.    </Topics>
16.  </ListTopicsResult>
17.  <ResponseMetadata>
18.    <RequestId></RequestId>
19.  </ResponseMetadata>
20. </ListTopicsResponse>

```

- if endpoint URL contain user/password information, in any of the topic, request must be made over HTTPS. If not, topic list request will be rejected.

## Notifications

Detailed under: [Bucket Operations](#).

Note

- “Abort Multipart Upload” request does not emit a notification
- Both “Initiate Multipart Upload” and “POST Object” requests will emit an `s3:ObjectCreated:Post` notification

## Events

The events are in JSON format (regardless of the actual endpoint), and share the same structure as the S3-compatible events pushed or pulled using the pubsub sync module. For example:

- ```

1. {"Records": [
2.   {

```

```

3.      "eventVersion":"2.1",
4.      "eventSource":"ceph:s3",
5.      "awsRegion":"us-east-1",
6.      "eventTime":"2019-11-22T13:47:35.124724Z",
7.      "eventName":"s3:ObjectCreated:Put",
8.      "userIdentity":{
9.          "principalId":"tester"
10.     },
11.     "requestParameters":{
12.         "sourceIPAddress": ""
13.     },
14.     "responseElements":{
15.         "x-amz-request-id":"503a4c37-85eb-47cd-8681-2817e80b4281.5330.903595",
16.         "x-amz-id-2":"14d2-zone1-zonegroup1"
17.     },
18.     "s3":{
19.         "s3SchemaVersion":"1.0",
20.         "configurationId":"mynotif1",
21.         "bucket": {
22.             "name": "mybucket1",
23.             "ownerIdentity": {
24.                 "principalId": "tester"
25.             },
26.             "arn": "arn:aws:s3:us-east-1::mybucket1",
27.             "id": "503a4c37-85eb-47cd-8681-2817e80b4281.5332.38"
28.         },
29.         "object": {
30.             "key": "myimage1.jpg",
31.             "size": "1024",
32.             "eTag": "37b51d194a7513e45b56f6524f2d51f2",
33.             "versionId": "",
34.             "sequencer": "F7E6D75DC742D108",
35.             "metadata": [],
36.             "tags": []
37.         }
38.     },
39.     "eventId": "",
40.     "opaqueData": "me@example.com"
41.   }
42. ]

```

- awsRegion: zonegroup
- eventTime: timestamp indicating when the event was triggered
- eventName: for list of supported events see: [S3 Notification Compatibility](#)
- userIdentity.principalId: user that triggered the change
- requestParameters.sourceIPAddress: not supported
- responseElements.x-amz-request-id: request ID of the original change

- `responseElements.x_amz_id_2`: RGW on which the change was made
- `s3.configurationId`: notification ID that created the event
- `s3.bucket.name`: name of the bucket
- `s3.bucket.ownerIdentity.principalId`: owner of the bucket
- `s3.bucket.arn`: ARN of the bucket
- `s3.bucket.id`: Id of the bucket (an extension to the S3 notification API)
- `s3.object.key`: object key
- `s3.object.size`: object size
- `s3.object.eTag`: object etag
- `s3.object.version`: object version in case of versioned bucket
- `s3.object.sequencer`: monotonically increasing identifier of the change per object (hexadecimal format)
- `s3.object.metadata`: any metadata set on the object sent as: `x-amz-meta-` (an extension to the S3 notification API)
- `s3.object.tags`: any tags set on the object (an extension to the S3 notification API)
- `s3.eventId`: unique ID of the event, that could be used for acking (an extension to the S3 notification API)
- `s3.opaqueData`: opaque data is set in the topic configuration and added to all notifications triggered by the topic (an extension to the S3 notification API)

# S3 Bucket Notifications Compatibility

Ceph's [Bucket Notifications](#) and [PubSub Module](#) APIs follow [AWS S3 Bucket Notifications API](#). However, some differences exist, as listed below.

## Note

Compatibility is different depending on which of the above mechanism is used

## Supported Destination

AWS supports: **SNS**, **SQS** and **Lambda** as possible destinations (AWS internal destinations). Currently, we support: **HTTP/S**, **Kafka** and **AMQP**. And also support pulling and acking of events stored in Ceph (as an internal destination).

We are using the **SNS** ARNs to represent the **HTTP/S**, **Kafka** and **AMQP** destinations.

## Notification Configuration XML

Following tags (and the tags inside them) are not supported:

Tag	Remarks
<QueueConfiguration>	not needed, we treat all destinations as SNS
<CloudFunctionConfiguration>	not needed, we treat all destinations as SNS

## REST API Extension

Ceph's bucket notification API has the following extensions:

- Deletion of a specific notification, or all notifications on a bucket, using the **DELETE** verb
  - In S3, all notifications are deleted when the bucket is deleted, or when an empty notification is set on the bucket
- Getting the information on a specific notification (when more than one exists on a bucket)
  - In S3, it is only possible to fetch all notifications on a bucket
- In addition to filtering based on prefix/suffix of object keys we support:
  - Filtering based on regular expression matching

- Filtering based on metadata attributes attached to the object
- Filtering based on object tags
- Each one of the additional filters extends the S3 API and using it will require extension of the client SDK (unless you are using plain HTTP).
- Filtering overlapping is allowed, so that same event could be sent as different notification

## Unsupported Fields in the Event Record

The records sent for bucket notification follow format described in: [Event Message Structure](#). However, the following fields may be sent empty, under the different deployment options (Notification/PubSub):

Field	Notification	PubSub	Description
<code>userIdentity.principalId</code>	Supported	Not Supported	The identity of the user that triggered the event
<code>requestParameters.sourceIPAddress</code>	Not Supported		The IP address of the client that triggered the event
<code>requestParameters.x-amz-request-id</code>	Supported	Not Supported	The request id that triggered the event
<code>requestParameters.x-amz-id-2</code>	Supported	Not Supported	The IP address of the RGW on which the event was triggered
<code>s3.object.size</code>	Supported	Not Supported	The size of the object

## Event Types

Event	Notification	PubSub
<code>s3:ObjectCreated:</code>	Supported	
<code>s3:ObjectCreated:Put</code>	Supported	Supported at <code>s3:ObjectCreated:</code> level
<code>s3:ObjectCreated:Post</code>	Supported	Not Supported
<code>s3:ObjectCreated:Copy</code>	Supported	Supported at <code>s3:ObjectCreated:</code> level

<code>s3:ObjectCreated:CompleteMultipartUpload</code>	Supported	Supported at <code>s3:ObjectCreated:</code> level
<code>s3:ObjectRemoved:*</code>	Supported	Supported only the specific events below
<code>s3:ObjectRemoved:Delete</code>	Supported	
<code>s3:ObjectRemoved:DeleteMarkerCreated</code>	Supported	
<code>s3:ObjectRestore:Post</code>		Not applicable to Ceph
<code>s3:ObjectRestore:Complete</code>		Not applicable to Ceph
<code>s3:ReducedRedundancyLostObject</code>		Not applicable to Ceph

## Topic Configuration

In the case of bucket notifications, the topics management API will be derived from [AWS Simple Notification Service API](#). Note that most of the API is not applicable to Ceph, and only the following actions are implemented:

- `CreateTopic`
- `DeleteTopic`
- `ListTopics`

We also have the following extensions to topic configuration:

- In `GetTopic` we allow fetching a specific topic, instead of all user topics
- In `CreateTopic`
  - we allow setting endpoint attributes
  - we allow setting opaque data that will be sent to the endpoint in the notification

# Rados Gateway Data Layout

Although the source code is the ultimate guide, this document helps new developers to get up to speed with the implementation details.

## Introduction

Swift offers something called a *container*, which we use interchangeably with the term *bucket*, so we say that RGW's buckets implement Swift containers.

This document does not consider how RGW operates on these structures, e.g. the use of encode() and decode() methods for serialization and so on.

## Conceptual View

Although RADOS only knows about pools and objects with their xattrs and omap[1], conceptually RGW organizes its data into three different kinds: metadata, bucket index, and data.

## Metadata

We have 3 'sections' of metadata: 'user', 'bucket', and 'bucket.instance'. You can use the following commands to introspect metadata entries:

```

1. $ radosgw-admin metadata list
2. $ radosgw-admin metadata list bucket
3. $ radosgw-admin metadata list bucket.instance
4. $ radosgw-admin metadata list user
5.
6. $ radosgw-admin metadata get bucket:<bucket>
7. $ radosgw-admin metadata get bucket.instance:<bucket>:<bucket_id>
8. $ radosgw-admin metadata get user:<user> # get or set

```

Some variables have been used in above commands, they are:

- user: Holds user information
- bucket: Holds a mapping between bucket name and bucket instance id
- bucket.instance: Holds bucket instance information[2]

Every metadata entry is kept on a single RADOS object. See below for implementation details.

Note that the metadata is not indexed. When listing a metadata section we do a RADOS `pgls` operation on the containing pool.

## Bucket Index

It's a different kind of metadata, and kept separately. The bucket index holds a key-value map in RADOS objects. By default it is a single RADOS object per bucket, but it is possible since Hammer to shard that map over multiple RADOS objects. The map itself is kept in omap, associated with each RADOS object. The key of each omap is the name of the objects, and the value holds some basic metadata of that object - metadata that shows up when listing the bucket. Also, each omap holds a header, and we keep some bucket accounting metadata in that header (number of objects, total size, etc.).

Note that we also hold other information in the bucket index, and it's kept in other key namespaces. We can hold the bucket index log there, and for versioned objects there is more information that we keep on other keys.

## Data

Objects data is kept in one or more RADOS objects for each rgw object.

## Object Lookup Path

---

When accessing objects, ReST APIs come to RGW with three parameters: account information (access key in S3 or account name in Swift), bucket or container name, and object name (or key). At present, RGW only uses account information to find out the user ID and for access control. Only the bucket name and object key are used to address the object in a pool.

The user ID in RGW is a string, typically the actual user name from the user credentials and not a hashed or mapped identifier.

When accessing a user's data, the user record is loaded from an object "<user\_id>" in pool "default.rgw.meta" with namespace "users.uid".

Bucket names are represented in the pool "default.rgw.meta" with namespace "root". Bucket record is loaded in order to obtain so-called marker, which serves as a bucket ID.

The object is located in pool "default.rgw.buckets.data". Object name is "<marker>\_<key>", for example "default.7593.4\_image.png", where the marker is "default.7593.4" and the key is "image.png". Since these concatenated names are not parsed, only passed down to RADOS, the choice of the separator is not important and causes no ambiguity. For the same reason, slashes are permitted in object names (keys).

It is also possible to create multiple data pools and make it so that different users buckets will be created in different RADOS pools by default, thus providing the necessary scaling. The layout and naming of these pools is controlled by a 'policy' setting.[3]

An RGW object may consist of several RADOS objects, the first of which is the head that contains the metadata, such as manifest, ACLs, content type, ETag, and user-defined metadata. The metadata is stored in xattrs. The head may also contain up to 512 kilobytes of object data, for efficiency and atomicity. The manifest describes how each object is laid out in RADOS objects.

## Bucket and Object Listing

---

Buckets that belong to a given user are listed in an omap of an object named “<user\_id>.buckets” (for example, “foo.buckets”) in pool “default.rgw.meta” with namespace “users.uid”. These objects are accessed when listing buckets, when updating bucket contents, and updating and retrieving bucket statistics (e.g. for quota).

See the user-visible, encoded class ‘cls\_user\_bucket\_entry’ and its nested class ‘cls\_user\_bucket’ for the values of these omap entries.

These listings are kept consistent with buckets in pool “.rgw”.

Objects that belong to a given bucket are listed in a bucket index, as discussed in sub-section ‘Bucket Index’ above. The default naming for index objects is “.dir.<marker>” in pool “default.rgw.buckets.index”.

## Footnotes

---

[1] Omap is a key-value store, associated with an object, in a way similar to how Extended Attributes associate with a POSIX file. An object’s omap is not physically located in the object’s storage, but its precise implementation is invisible and immaterial to RADOS Gateway. In Hammer, one LevelDB is used to store omap in each OSD.

[2] Before the Dumpling release, the ‘bucket.instance’ metadata did not exist and the ‘bucket’ metadata contained its information. It is possible to encounter such buckets in old installations.

[3] The pool names have been changed starting with the Infernalis release. If you are looking at an older setup, some details may be different. In particular there was a different pool for each of the namespaces that are now being used inside the default.root.meta pool.

## Appendix: Compendium

---

Known pools:

.rgw.root

Unspecified region, zone, and global information records, one per object.

<zone>.rgw.control

```
notify.<N>
```

```
<zone>.rgw.meta
```

Multiple namespaces with different kinds of metadata:

- namespace: root

```
<bucket> .bucket.meta.<bucket>:<marker> # see put_bucket_instance_info()
```

The tenant is used to disambiguate buckets, but not bucket instances. Example:

```
1. .bucket.meta.prodtx:test%25star:default.84099.6
2. .bucket.meta.testcont:default.4126.1
3. .bucket.meta.prodtx:testcont:default.84099.4
4. prodtx/testcont
5. prodtx/test%25star
6. testcont
```

namespace: users.uid

Contains \_both\_ per-user information (RGWUserInfo) in "<user>" objects and per-user lists of buckets in ommaps of "<user>.buckets" objects. The "<user>" may contain the tenant if non-empty, for example:

```
1. prodtx$prod
2. test2.buckets
3. prodtx$prod.buckets
4. test2
```

namespace: users.email

Unimportant

namespace: users.keys

47UA98JSTJZ9YAN30S30

This allows `radosgw` to look up users by their access keys during authentication.

namespace: users.swift

test:tester

```
<zone>.rgw.buckets.index
```

Objects are named ".dir.<marker>", each contains a bucket index. If the index is sharded, each shard appends the shard index after the marker.

```
<zone>.rgw.buckets.data
```

```
default.7593.4__shadow_.488urDFerTYXavx4yAd-0p8mxehnvTI_1 <marker>_<key>
```

An example of a marker would be "default.16004.1" or "default.7593.4". The current format is "<zone>.<instance\_id>.<bucket\_id>". But once generated, a marker is not parsed again, so its format may change freely in the future.

# STS in Ceph

---

Secure Token Service is a web service in AWS that returns a set of temporary security credentials for authenticating federated users. The link to official AWS documentation can be found here: <https://docs.aws.amazon.com/STS/latest/APIReference>Welcome.html>.

Ceph Object Gateway implements a subset of STS APIs that provide temporary credentials for identity and access management. These temporary credentials can be used to make subsequent S3 calls which will be authenticated by the STS engine in Ceph Object Gateway. Permissions of the temporary credentials can be further restricted via an IAM policy passed as a parameter to the STS APIs.

## STS REST APIs

---

The following STS REST APIs have been implemented in Ceph Object Gateway:

1. **AssumeRole**: Returns a set of temporary credentials that can be used for cross-account access. The temporary credentials will have permissions that are allowed by both - permission policies attached with the Role and policy attached with the AssumeRole API.

Parameters:

**RoleArn** (String/ Required): ARN of the Role to Assume.

**RoleSessionName** (String/ Required): An Identifier for the assumed role session.

**Policy** (String/ Optional): An IAM Policy in JSON format.

**DurationSeconds** (Integer/ Optional): The duration in seconds of the session. Its default value is 3600.

**ExternalId** (String/ Optional): A unique Id that might be used when a role is assumed in another account.

**SerialNumber** (String/ Optional): The Id number of the MFA device associated with the user making the AssumeRole call.

**TokenCode** (String/ Optional): The value provided by the MFA device, if the trust policy of the role being assumed requires MFA.

2. **AssumeRoleWithWebIdentity**: Returns a set of temporary credentials for users that have been authenticated by a web/mobile app by an OpenID Connect /OAuth2.0 Identity Provider. Currently Keycloak has been tested and integrated with RGW.

Parameters:

**RoleArn** (String/ Required): ARN of the Role to Assume.

**RoleSessionName** (String/ Required): An Identifier for the assumed role session.

**Policy** (String/ Optional): An IAM Policy in JSON format.

**DurationSeconds** (Integer/ Optional): The duration in seconds of the session. Its default value is 3600.

**ProviderId** (String/ Optional): Fully qualified host component of the domain name of the IDP. Valid only for OAuth2.0 tokens (not for OpenID Connect tokens).

**WebIdentityToken** (String/ Required): The OpenID Connect/ OAuth2.0 token, which the application gets in return after authenticating its user with an IDP.

Before invoking `AssumeRoleWithWebIdentity`, an OpenID Connect Provider entity (which the web application authenticates with), needs to be created in RGW.

The trust between the IDP and the role is created by adding a Condition to the role trust policy, which allows access only to applications with the app id given in the trust policy document. The Condition is of the form:

```
    {"\"Version\": \"2012-10-17\", \"Statement\":[{\"\"Effect\": \"Allow\", \"Principal\":[\"Federated\": [\"arn:aws:iam:::oidc-provider/<URL of IDP>\"]], \"Action\":[\"sts:AssumeRoleWithWebIdentity\"], \"Condition\": {\"StringEquals\": {\"<URL of IDP> :app_id\": \"<aud>\"}}}], \"Condition\": {\"StringEquals\": {\"<URL of IDP> :app_id\": \"<aud>\"}}}]}
```

The `app_id` in the condition above must match the 'aud' field of the incoming token.

A shadow user is created corresponding to every federated user. The user id is derived from the 'sub' field of the incoming web token. The user is created in a separate namespace - 'oidc' such that the user id doesn't clash with any other user ids in rgw. The format of the user id is - <tenant>\$<user-namespace>\$<sub> where user-namespace is 'oidc' for users that authenticate with oidc providers.

## STS Configuration

The following configurable options have to be added for STS integration:

```
1. [client.radosgw.gateway]
2. rgw sts key = {sts key for encrypting the session token}
3. rgw s3 auth use_sts = true
```

Note: By default, STS and S3 APIs co-exist in the same namespace, and both S3 and STS APIs can be accessed via the same endpoint in Ceph Object Gateway.

## Examples

1. The following is an example of AssumeRole API call, which shows steps to create a role, assign a policy to it (that allows access to S3 resources), assuming a role to get temporary credentials and accessing s3 resources using those credentials. In this

example, TESTER1 assumes a role created by TESTER, to access S3 resources owned by TESTER, according to the permission policy attached to the role.

```
1. import boto3
2.
3. iam_client = boto3.client('iam',
4. aws_access_key_id=<access_key of TESTER>,
5. aws_secret_access_key=<secret_key of TESTER>,
6. endpoint_url=<IAM URL>,
7. region_name=''
8. )
9.
10. policy_document = "{\"Version\":\"2012-10-17\", \"Statement\":[{\"Effect\":\"Allow\", \"Principal\":\"AWS\":
11. [\"arn:aws:iam::user/TESTER1\"]}, {"Action\":[\"sts:AssumeRole\"]}]}"
12. role_response = iam_client.create_role(
13. AssumeRolePolicyDocument=policy_document,
14. Path='/',
15. RoleName='S3Access',
16. )
17.
18. role_policy = "{\"Version\":\"2012-10-17\", \"Statement\":
19. {\"Effect\":\"Allow\", \"Action\":[\"s3:*\"], \"Resource\":[\"arn:aws:s3:::*\"]}}"
20. response = iam_client.put_role_policy(
21. RoleName='S3Access',
22. PolicyName='Policy1',
23. PolicyDocument=role_policy
24. )
25.
26. sts_client = boto3.client('sts',
27. aws_access_key_id=<access_key of TESTER1>,
28. aws_secret_access_key=<secret_key of TESTER1>,
29. endpoint_url=<STS URL>,
30. region_name='',
31. )
32.
33. response = sts_client.assume_role(
34. RoleArn=role_response['Role']['Arn'],
35. RoleSessionName='Bob',
36. DurationSeconds=3600
37. )
38.
39. s3client = boto3.client('s3',
40. aws_access_key_id = response['Credentials']['AccessKeyId'],
41. aws_secret_access_key = response['Credentials']['SecretAccessKey'],
42. aws_session_token = response['Credentials']['SessionToken'],
43. endpoint_url=<S3 URL>,
44. region_name='', )
45.
46. bucket_name = 'my-bucket'
47. s3bucket = s3client.create_bucket(Bucket=bucket_name)
48. resp = s3client.list_buckets()
```

2. The following is an example of AssumeRoleWithWebIdentity API call, where an external app that has users authenticated with an OpenID Connect/ OAuth2 IDP (Keycloak in this example), assumes a role to get back temporary credentials and access S3 resources according to permission policy of the role.

```
1. import boto3
2.
3. iam_client = boto3.client('iam',
4.     aws_access_key_id=<access_key of TESTER>,
5.     aws_secret_access_key=<secret_key of TESTER>,
6.     endpoint_url=<IAM URL>,
7.     region_name=''
8. )
9.
10. oidc_response = iam_client.create_open_id_connect_provider(
11.     Url=<URL of the OpenID Connect Provider>,
12.     ClientIDList=[  

13.         <Client id registered with the IDP>
14.     ],
15.     ThumbprintList=[  

16.         <Thumbprint of the IDP>
17. ]
18. )
19.
20. policy_document = "{\"Version\":\"2012-10-17\", \"Statement\":[{\"Effect\":\"Allow\", \"Principal\":  

21.     \"Federated\":[\"arn:aws:iam::oidc-provider/localhost:8080/auth/realms/demo\"], \"Action\":  

22.     [\"sts:AssumeRoleWithWebIdentity\"]}, {\"Condition\":{\"StringEquals\":  

23.         {\"localhost:8080/auth/realms/demo:app_id\":\"customer-portal\"}}}]}"
24. role_response = iam_client.create_role(  

25.     AssumeRolePolicyDocument=policy_document,  

26.     Path='/',
27.     RoleName='S3Access',
28. )
29. role_policy = "{\"Version\":\"2012-10-17\", \"Statement\":  

30.     {\"Effect\":\"Allow\", \"Action\":\"s3:*\", \"Resource\":\"arn:aws:s3:::*\"}}"
31. response = iam_client.put_role_policy(  

32.     RoleName='S3Access',
33.     PolicyName='Policy1',
34.     PolicyDocument=role_policy
35. )
36. sts_client = boto3.client('sts',
37.     aws_access_key_id=<access_key of TESTER1>,
38.     aws_secret_access_key=<secret_key of TESTER1>,
39.     endpoint_url=<STS URL>,
40.     region_name='',
41. )
42. response = client.assume_role_with_web_identity(  

43.     RoleArn=role_response['Role']['Arn'],
```

```

44. RoleSessionName='Bob',
45. DurationSeconds=3600,
46. WebIdentityToken=<Web Token>
47. )
48.

49. s3client = boto3.client('s3',
50. aws_access_key_id = response['Credentials']['AccessKeyId'],
51. aws_secret_access_key = response['Credentials']['SecretAccessKey'],
52. aws_session_token = response['Credentials']['SessionToken'],
53. endpoint_url=<S3 URL>,
54. region_name=''),
55.

56. bucket_name = 'my-bucket'
57. s3bucket = s3client.create_bucket(Bucket=bucket_name)
58. resp = s3client.list_buckets()

```

## How to obtain thumbprint of an OpenID Connect Provider IDP

1. Take the OpenID connect provider's URL and add /.well-known/openid-configuration to it to get the URL to get the IDP's configuration document. For example, if the URL of the IDP is <http://localhost:8000/auth/realm/quickstart>, then the URL to get the document from is <http://localhost:8000/auth/realm/quickstart/.well-known/openid-configuration>
2. Use the following curl command to get the configuration document from the URL described in step 1:

```

1. curl -k -v \
2.   -X GET \
3.   -H "Content-Type: application/x-www-form-urlencoded" \
4.   "http://localhost:8000/auth/realm/quickstart/.well-known/openid-configuration" \
5.   | jq .
6.

7. 3. From the response of step 2, use the value of "jwks_uri" to get the certificate of the IDP,
8. using the following code::
9.   curl -k -v \
10.    -X GET \
11.    -H "Content-Type: application/x-www-form-urlencoded" \
12.    "http://$KC_SERVER/$KC_CONTEXT/realm/$KC_REALM/protocol/openid-connect/certs" \
13.    | jq .

```

3. Copy the result of "x5c" in the response above, in a file certificate.crt, and add '--BEGIN CERTIFICATE--' at the beginning and "--END CERTIFICATE--" at the end.

1. Use the following OpenSSL command to get the certificate thumbprint:

```
1. openssl x509 -in certificate.crt -fingerprint -noout
```

2. The result of the above command in step 4, will be a SHA1 fingerprint, like the following:

```
1. SHA1 Fingerprint=F7:D7:B3:51:5D:D0:D3:19:DD:21:9A:43:A9:EA:72:7A:D6:06:52:87
```

6. Remove the colons from the result above to get the final thumbprint which can be as input while creating the OpenID Connect Provider entity in IAM:

```
1. F7D7B3515DD0D319DD219A43A9EA727AD6065287
```

## Roles in RGW

More information for role manipulation can be found here [Role](#).

## OpenID Connect Provider in RGW

More information for OpenID Connect Provider entity manipulation can be found here [OpenID Connect Provider in RGW](#).

## Keycloak integration with Radosgw

Steps for integrating Radosgw with Keycloak can be found here [Keycloak integration with RadosGW](#).

## STSLite

STSLite has been built on STS, and documentation for the same can be found here [STS Lite](#).

# STS Lite

Ceph Object Gateway provides support for a subset of Amazon Secure Token Service (STS) APIs. STS Lite is an extension of STS and builds upon one of its APIs to decrease the load on external IDPs like Keystone and LDAP.

A set of temporary security credentials is returned after authenticating a set of AWS credentials with the external IDP. These temporary credentials can be used to make subsequent S3 calls which will be authenticated by the STS engine in Ceph, resulting in less load on the Keystone/ LDAP server.

Temporary and limited privileged credentials can be obtained for a local user also using the STS Lite API.

## STS Lite REST APIs

The following STS Lite REST API is part of STS Lite in Ceph Object Gateway:

**1. GetSessionToken:** Returns a set of temporary credentials for a set of AWS credentials. After initial authentication with Keystone/ LDAP, the temporary credentials returned can be used to make subsequent S3 calls. The temporary credentials will have the same permission as that of the AWS credentials.

Parameters:

**DurationSeconds** (Integer/ Optional): The duration in seconds for which the credentials should remain valid. Its default value is 3600. Its default max value is 43200 which is can be configured using rgw sts max session duration.

**SerialNumber** (String/ Optional): The Id number of the MFA device associated with the user making the GetSessionToken call.

**TokenCode** (String/ Optional): The value provided by the MFA device, if MFA is required.

An administrative user needs to attach a policy to allow invocation of GetSessionToken API using its permanent credentials and to allow subsequent s3 operations invocation using only the temporary credentials returned by GetSessionToken.

The user attaching the policy needs to have admin caps. For example:

```
1. radosgw-admin caps add --uid="TESTER" --caps="user-policy=*" 
```

The following is the policy that needs to be attached to a user 'TESTER1':

```

user_policy = "{\"Version\":\"2012-10-17\", \"Statement\":
[{\\"Effect\\\":\"Deny\", \\"Action\\\":\"s3:*\", \\"Resource\\\":[\"*\"], \\"Condition\\\":{\\\"BoolIfExists\\\":
\\\"sts:authentication\\\":\\\"false\\\"}}},
{\\"Effect\\\":\"Allow\", \\"Action\\\\":\"sts:GetSessionToken\", \\"Resource\\\\":\"*\", \\"Condition\\\\":{\\\"BoolIfExists\\\":
1. {\\\"sts:authentication\\\":\\\"false\\\"}}}]}"

```

## STS Lite Configuration

The following configurable options are available for STS Lite integration:

1. [client.radosgw.gateway]
2. rgw sts key = {sts key for encrypting the session token}
3. rgw s3 auth use sts = true

The above STS configurables can be used with the Keystone configurables if one needs to use STS Lite in conjunction with Keystone. The complete set of configurable options will be:

1. [client.radosgw.gateway]
2. rgw sts key = {sts key for encrypting/ decrypting the session token}
3. rgw s3 auth use sts = true
- 4.
5. rgw keystone url = {keystone server url:keystone server admin port}
6. rgw keystone admin project = {keystone admin project name}
7. rgw keystone admin tenant = {keystone service tenant name}
8. rgw keystone admin domain = {keystone admin domain name}
9. rgw keystone api version = {keystone api version}
10. rgw keystone implicit tenants = {true for private tenant for each new user}
11. rgw keystone admin password = {keystone service tenant user name}
12. rgw keystone admin user = keystone service tenant user password}
13. rgw keystone accepted roles = {accepted user roles}
14. rgw keystone token cache size = {number of tokens to cache}
15. rgw s3 auth use keystone = true

The details of the integrating ldap with Ceph Object Gateway can be found here:

[Integrating with OpenStack Keystone](#)

The complete set of configurables to use STS Lite with LDAP are:

1. [client.radosgw.gateway]
2. rgw sts key = {sts key for encrypting/ decrypting the session token}
3. rgw s3 auth use sts = true
- 4.
5. rgw\_s3\_auth\_use\_ldap = true
6. rgw\_ldap\_uri = {LDAP server to use}
7. rgw\_ldap\_binddn = {Distinguished Name (DN) of the service account}
8. rgw\_ldap\_secret = {password for the service account}
9. rgw\_ldap\_searchdn = {base in the directory information tree for searching users}
10. rgw\_ldap\_dnattr = {attribute being used in the constructed search filter to match a username}

```
11. rgw_ldap_searchfilter = {search filter}
```

The details of the integrating ldap with Ceph Object Gateway can be found here: [LDAP Authentication](#)

Note: By default, STS and S3 APIs co-exist in the same namespace, and both S3 and STS APIs can be accessed via the same endpoint in Ceph Object Gateway.

## Example showing how to Use STS Lite with Keystone

The following are the steps needed to use STS Lite with Keystone. Boto 3.x has been used to write an example code to show the integration of STS Lite with Keystone.

### 1. Generate EC2 credentials :

```
1. openstack ec2 credentials create
2. +-----+-----+
3. | Field      | Value
4. +-----+-----+
5. | access      | b924dfc87d454d15896691182fdeb0ef
6. | links       | {u'self': u'http://192.168.0.15/identity/v3/users/'
7. |             | 40a7140e424f493d8165abc652dc731c/credentials/
8. |             | OS-EC2/b924dfc87d454d15896691182fdeb0ef'}
9. | project_id   | c703801dccaf4a0aaa39bec8c481e25a
10. | secret       | 6a2142613c504c42a94ba2b82147dc28
11. | trust_id     | None
12. | user_id      | 40a7140e424f493d8165abc652dc731c
13. +-----+-----+
```

### 1. Use the credentials created in the step 1. to get back a set of temporary credentials using GetSessionToken API.

```
1. import boto3
2.
3. access_key = <ec2 access key>
4. secret_key = <ec2 secret key>
5.
6. client = boto3.client('sts',
7. aws_access_key_id=access_key,
8. aws_secret_access_key=secret_key,
9. endpoint_url=<STS URL>,
10. region_name='',
11. )
12.
13. response = client.get_session_token(
14.     DurationSeconds=43200
15. )
```

1. The temporary credentials obtained in step 2. can be used for making S3 calls:

```

1. s3client = boto3.client('s3',
2.     aws_access_key_id = response['Credentials']['AccessKeyId'],
3.     aws_secret_access_key = response['Credentials']['SecretAccessKey'],
4.     aws_session_token = response['Credentials']['SessionToken'],
5.     endpoint_url=<S3 URL>,
6.     region_name=' ')
7.
8. bucket = s3client.create_bucket(Bucket='my-new-shiny-bucket')
9. response = s3client.list_buckets()
10. for bucket in response["Buckets"]:
11.     print "{name}\t{created}".format(
12.         name = bucket['Name'],
13.         created = bucket['CreationDate'],
14.     )

```

Similar steps can be performed for using GetSessionToken with LDAP.

## Limitations and Workarounds

1. Keystone currently supports only S3 requests, hence in order to successfully authenticate an STS request, the following workaround needs to be added to boto to the following file - botocore/auth.py

Lines 13-16 have been added as a workaround in the code block below:

```

1. class SigV4Auth(BaseSigner):
2.     """
3.     Sign a request with Signature V4.
4.     """
5.     REQUIRES_REGION = True
6.
7.     def __init__(self, credentials, service_name, region_name):
8.         self.credentials = credentials
9.         # We initialize these value here so the unit tests can have
10.        # valid values. But these will get overridden in ``add_auth``
11.        # later for real requests.
12.        self._region_name = region_name
13.        if service_name == 'sts':
14.            self._service_name = 's3'
15.        else:
16.            self._service_name = service_name

```

# Keycloak integration with RadosGW

Keycloak can be setup as an OpenID Connect Identity Provider, which can be used by mobile/ web apps to authenticate their users. The Web token returned as a result of authentication can be used by the mobile/ web app to call AssumeRoleWithWebIdentity to get back a set of temporary S3 credentials, which can be used by the app to make S3 calls.

## Setting up Keycloak

Installing and bringing up Keycloak can be found here:  
[https://www.keycloak.org/docs/latest/server\\_installation/](https://www.keycloak.org/docs/latest/server_installation/).

## Configuring Keycloak to talk to RGW

The following configurables have to be added for RGW to talk to Keycloak. The format of token inspection url is [https://\[base-server-url\]/token/introspect](https://[base-server-url]/token/introspect):

1. [client.radosgw.gateway]
2. rgw sts key = {sts key for encrypting/ decrypting the session token}
3. rgw s3 auth use sts = true

## Example showing how to fetch a web token from Keycloak

Several examples of apps authenticating with Keycloak are given here:

<https://github.com/keycloak/keycloak-quickstarts/blob/latest/docs/getting-started.md>

Taking the example of app-profile-jee-jsp app given in the link above, its client secret and client password, can be used to fetch the access token (web token) as given below:

1. KC\_REALM=demo
2. KC\_CLIENT=<client id>
3. KC\_CLIENT\_SECRET=<client secret>
4. KC\_SERVER=<host>:8080
5. KC\_CONTEXT=auth
- 6.
7. # Request Tokens for credentials
8. KC\_RESPONSE=\$( \
9. curl -k -v -X POST \
10. -H "Content-Type: application/x-www-form-urlencoded" \
11. -d "scope=openid" \
12. -d "grant\_type=client\_credentials" \
13. -d "client\_id=\$KC\_CLIENT" \

```
14. -d "client_secret=$KC_CLIENT_SECRET" \
15. "http://$KC_SERVER/$KC_CONTEXT/realms/$KC_REALM/protocol/openid-connect/token" \
16. | jq .
17. )
18.
19. KC_ACCESS_TOKEN=$(echo $KC_RESPONSE| jq -r .access_token)
```

KC\_ACCESS\_TOKEN can be used to invoke AssumeRoleWithWebIdentity as given in [STS in Ceph](#).

# Role

A role is similar to a user and has permission policies attached to it, that determine what a role can or can not do. A role can be assumed by any identity that needs it. If a user assumes a role, a set of dynamically created temporary credentials are returned to the user. A role can be used to delegate access to users, applications, services that do not have permissions to access some s3 resources.

The following radosgw-admin commands can be used to create/ delete/ update a role and permissions associated with a role.

## Create a Role

To create a role, execute the following:

```
radosgw-admin role create --role-name={role-name} [--path=="{path to the role}"] [--assume-role-policy-doc=1. {trust-policy-document}]
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

`path`

Description

Path to the role. The default value is a slash(/).

Type

String

`assume-role-policy-doc`

Description

The trust relationship policy document that grants an entity permission to assume the role.

Type

String

For example:

```
1. radosgw-admin role create --role-name=S3Access1 --path=/application_abc/component_xyz/ --assume-role-policy-doc=\{"Version\":\"2012-10-17\",\"Statement\":[{\\"Effect\":\"Allow\",\"Principal\":\\\"AWS\\\":\\\"arn:aws:iam::user/TESTER\\\"},\\"Action\":[\\\"sts:AssumeRole\\\"]}\\}
```

```
1. {
2.   "id": "ca43045c-082c-491a-8af1-2eebca13deec",
3.   "name": "S3Access1",
4.   "path": "/application_abc/component_xyz/",
5.   "arn": "arn:aws:iam::::role/application_abc/component_xyz/S3Access1",
6.   "create_date": "2018-10-17T10:18:29.116Z",
7.   "max_session_duration": 3600,
8.   "assume_role_policy_document": "{\"Version\":\"2012-10-17\",\"Statement\":[{\\"Effect\":\"Allow\",\"Principal\":\\\"AWS\\\":\\\"arn:aws:iam::user/TESTER\\\"},\\"Action\":[\\\"sts:AssumeRole\\\"]}]"
9. }
```

## Delete a Role

To delete a role, execute the following:

```
1. radosgw-admin role rm --role-name={role-name}
```

## Request Parameters

**role-name**

Description

Name of the role.

Type

String

For example:

```
1. radosgw-admin role rm --role-name=S3Access1
```

Note: A role can be deleted only when it doesn't have any permission policy attached to it.

## Get a Role

To get information about a role, execute the following:

```
1. radosgw-admin role get --role-name={role-name}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

For example:

```
1. radosgw-admin role get --role-name=S3Access1
```

```
1. {
2.   "id": "ca43045c-082c-491a-8af1-2eebca13deec",
3.   "name": "S3Access1",
4.   "path": "/application_abc/component_xyz/",
5.   "arn": "arn:aws:iam::::role/application_abc/component_xyz/S3Access1",
6.   "create_date": "2018-10-17T10:18:29.116Z",
7.   "max_session_duration": 3600,
8.   "assume_role_policy_document": "{\"Version\":\"2012-10-17\",\"Statement\":
[{\\"Effect\":\"Allow\",\"Principal\":{\"AWS\":[\"arn:aws:iam::user/TESTER\"]},\"Action\":
[\"sts:AssumeRole\"]}]}"
9. }
```

## List Roles

To list roles with a specified path prefix, execute the following:

```
1. radosgw-admin role list [--path-prefix ={path prefix}]
```

## Request Parameters

`path-prefix`

Description

Path prefix for filtering roles. If this is not specified, all roles are listed.

Type

String

For example:

```
1. radosgw-admin role list --path-prefix="/application"

1. [
2.   {
3.     "id": "3e1c0ff7-8f2b-456c-8fdf-20f428ba6a7f",
4.     "name": "S3Access1",
5.     "path": "/application_abc/component_xyz/",
6.     "arn": "arn:aws:iam:::role/application_abc/component_xyz/S3Access1",
7.     "create_date": "2018-10-17T10:32:01.881Z",
8.     "max_session_duration": 3600,
9.     "assume_role_policy_document": "{\"Version\":\"2012-10-17\", \"Statement\": [{\"Effect\":\"Allow\", \"Principal\":{\"AWS\":[\"arn:aws:iam::user/TESTER\"]}, \"Action\": [\"sts:AssumeRole\"]}]}"
10.   }
11. ]
```

## Update Assume Role Policy Document of a role

To modify a role's assume role policy document, execute the following:

```
1. radosgw-admin role modify --role-name={role-name} --assume-role-policy-doc={trust-policy-document}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

`assume-role-policy-doc`

Description

The trust relationship policy document that grants an entity permission to assume the role.

Type

String

For example:

```
radosgw-admin role modify --role-name=S3Access1 --assume-role-policy-doc={"Version":"2012-10-17","Statement":[{"Effect":"Allow","Principal":["AWS":["arn:aws:iam::user/TESTER2"]],"Action":["sts:AssumeRole"]}]}
```

```
1. {
2.   "id": "ca43045c-082c-491a-8af1-2eebca13deec",
3.   "name": "S3Access1",
4.   "path": "/application_abc/component_xyz/",
5.   "arn": "arn:aws:iam::role/application_abc/component_xyz/S3Access1",
6.   "create_date": "2018-10-17T10:18:29.116Z",
7.   "max_session_duration": 3600,
8.   "assume_role_policy_document": "{\"Version\":\"2012-10-17\",\"Statement\": [{\"Effect\":\"Allow\",\"Principal\":{\"AWS\":["arn:aws:iam::user/TESTER2"]},\"Action\": [\"sts:AssumeRole\"]}]}"
9. }
```

In the above example, we are modifying the Principal from TESTER to TESTER2 in its assume role policy document.

## Add/ Update a Policy attached to a Role

To add or update the inline policy attached to a role, execute the following:

```
radosgw-admin role policy put --role-name={role-name} --policy-name={policy-name} --policy-doc={permission-policy-doc}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

`policy-name`

Description

Name of the policy.

Type

String

`policy-doc`

Description

Role

The Permission policy document.

Type

String

For example:

```
radosgw-admin role-policy put --role-name=S3Access1 --policy-name=Policy1 --policy-doc=\{"Version\":\"2012-10-17\", \"Statement\":[{\\"Effect\":\"Allow\", \\"Action\":\"\\s3:*\"}, {\\"Resource\":\"arn:aws:s3:::example_bucket\"}]\}
```

In the above example, we are attaching a policy 'Policy1' to role 'S3Access1', which allows all s3 actions on 'example\_bucket'.

## List Permission Policy Names attached to a Role

To list the names of permission policies attached to a role, execute the following:

```
1. radosgw-admin role policy get --role-name={role-name}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

For example:

```
1. radosgw-admin role-policy list --role-name=S3Access1
```

```
1. [  
2.   "Policy1"  
3. ]
```

## Get Permission Policy attached to a Role

To get a specific permission policy attached to a role, execute the following:

```
1. radosgw-admin role policy get --role-name={role-name} --policy-name={policy-name}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

String

`policy-name`

Description

Name of the policy.

Type

String

For example:

```
1. radosgw-admin role-policy get --role-name=S3Access1 --policy-name=Policy1
```

```
1. {
    "Permission policy": "{\"Version\":\"2012-10-17\",\"Statement\":[{\"Effect\":\"Allow\",\"Action\":
2. [\"s3:*\"],\"Resource\":\"arn:aws:s3:::example_bucket\"}]}"
3. }
```

## Delete Policy attached to a Role

To delete permission policy attached to a role, execute the following:

```
1. radosgw-admin role policy rm --role-name={role-name} --policy-name={policy-name}
```

## Request Parameters

`role-name`

Description

Name of the role.

Type

## String

```
policy-name
```

### Description

Name of the policy.

### Type

## String

For example:

```
1. radosgw-admin role-policy get --role-name=S3Access1 --policy-name=Policy1
```

## REST APIs for Manipulating a Role

In addition to the above radosgw-admin commands, the following REST APIs can be used for manipulating a role. For the request parameters and their explanations, refer to the sections above.

In order to invoke the REST admin APIs, a user with admin caps needs to be created.

```
1. radosgw-admin --uid TESTER --display-name "TestUser" --access_key TESTER --secret test123 user create
2. radosgw-admin caps add --uid="TESTER" --caps="roles=*"
```

## Create a Role

Example::

```
POST "<hostname>?
Action=CreateRole&RoleName=S3Access&Path=/application_abc/component_xyz/&AssumeRolePolicyDocument={"Version":"2012-10-17","Statement":[{"Effect":"Allow","Principal":{"AWS":["arn:aws:iam:::user/TESTER"]},"Action":["sts:AssumeRole"]}]}"
```

```
1. <role>
2. <id>8f41f4e0-7094-4dc0-ac20-074a881ccbc5</id>
3. <name>S3Access</name>
4. <path>/application_abc/component_xyz/</path>
5. <arn>arn:aws:iam::role/application_abc/component_xyz/S3Access</arn>
6. <create_date>2018-10-23T07:43:42.811Z</create_date>
7. <max_session_duration>3600</max_session_duration>
   <assume_role_policy_document>{"Version":"2012-10-17","Statement":[{"Effect":"Allow","Principal":{"AWS":["arn:aws:iam:::user/TESTER"]},"Action":["sts:AssumeRole"]}]}</assume_role_policy_document>
8. </role>
```

## Delete a Role

Example::

```
POST "<hostname>?Action=DeleteRole&RoleName=S3Access"
```

Note: A role can be deleted only when it doesn't have any permission policy attached to it.

## Get a Role

Example::

```
POST "<hostname>?Action=GetRole&RoleName=S3Access"
```

```

1. <role>
2.   <id>8f41f4e0-7094-4dc0-ac20-074a881ccbc5</id>
3.   <name>S3Access</name>
4.   <path>/application_abc/component_xyz/</path>
5.   <arn>arn:aws:iam:::role/application_abc/component_xyz/S3Access</arn>
6.   <create_date>2018-10-23T07:43:42.811Z</create_date>
7.   <max_session_duration>3600</max_session_duration>
     <assume_role_policy_document>{"Version":"2012-10-17", "Statement": [{"Effect":"Allow", "Principal":{"AWS":
8.   ["arn:aws:iam:::user/TESTER"]}, "Action":["sts:AssumeRole"]}]}</assume_role_policy_document>
9. </role>
```

## List Roles

Example::

```
POST "<hostname>?Action=ListRoles&RoleName=S3Access&PathPrefix=/application"
```

```

1. <role>
2.   <id>8f41f4e0-7094-4dc0-ac20-074a881ccbc5</id>
3.   <name>S3Access</name>
4.   <path>/application_abc/component_xyz/</path>
5.   <arn>arn:aws:iam:::role/application_abc/component_xyz/S3Access</arn>
6.   <create_date>2018-10-23T07:43:42.811Z</create_date>
7.   <max_session_duration>3600</max_session_duration>
     <assume_role_policy_document>{"Version":"2012-10-17", "Statement": [{"Effect":"Allow", "Principal":{"AWS":
8.   ["arn:aws:iam:::user/TESTER"]}, "Action":["sts:AssumeRole"]}]}</assume_role_policy_document>
9. </role>
```

## Update Assume Role Policy Document

Example::

```
POST "<hostname>?Action=UpdateAssumeRolePolicy&RoleName=S3Access&PolicyDocument=
{"Version":"2012-10-17", "Statement": [{"Effect":"Allow", "Principal":{"AWS":
["arn:aws:iam:::user/TESTER2"]}, "Action":["sts:AssumeRole"]}]}"
```

## Add/ Update a Policy attached to a Role

Example::

```
POST "<hostname>?  
Action=PutRolePolicy&RoleName=S3Access&PolicyName=Policy1&PolicyDocument=  
{"Version":"2012-10-17","Statement":[{"Effect":"Allow","Action":  
["s3>CreateBucket"],"Resource":"arn:aws:s3:::example_bucket"}]}"
```

## List Permission Policy Names attached to a Role

Example::

```
POST "<hostname>?Action=ListRolePolicies&RoleName=S3Access"
```

```
1. <PolicyNames>  
2. <member>Policy1</member>  
3. </PolicyNames>
```

## Get Permission Policy attached to a Role

Example::

```
POST "<hostname>?Action=GetRolePolicy&RoleName=S3Access&PolicyName=Policy1"
```

```
1. <GetRolePolicyResult>  
2. <PolicyName>Policy1</PolicyName>  
3. <RoleName>S3Access</RoleName>  
4. <Permission_policy>{"Version":"2012-10-17","Statement":[{"Effect":"Allow","Action":  
["s3>CreateBucket"],"Resource":"arn:aws:s3:::example_bucket"}]}</Permission_policy>  
5. </GetRolePolicyResult>
```

## Delete Policy attached to a Role

Example::

```
POST "<hostname>?Action=DeleteRolePolicy&RoleName=S3Access&PolicyName=Policy1"
```

# Orphan List and Associated Tooling

## Contents

- [Orphan List and Associated Tooling](#)
  - [Orphans Find – DEPRECATED](#)
  - [Orphan List](#)
    - [WARNING: Experimental Status](#)
    - [WARNING: Specifying a Data Pool](#)
    - [WARNING: Unindexed Buckets](#)
  - [RADOS List](#)
    - [Note: Shared Bucket Markers](#)

Orphans are RADOS objects that are left behind after their associated RGW objects are removed. Normally these RADOS objects are removed automatically, either immediately or through a process known as “garbage collection”. Over the history of RGW, however, there may have been bugs that prevented these RADOS objects from being deleted, and these RADOS objects may be consuming space on the Ceph cluster without being of any use. From the perspective of RGW, we call such RADOS objects “orphans”.

## Orphans Find – DEPRECATED

The radosgw-admin tool has/had three subcommands to help manage orphans, however these subcommands are (or will soon be) deprecated. These subcommands are:

::

```
# radosgw-admin orphans find ... # radosgw-admin orphans finish ... # radosgw-admin  
orphans list-jobs ...
```

There are two key problems with these subcommands, however. First, these subcommands have not been actively maintained and therefore have not tracked RGW as it has evolved in terms of features and updates. As a result the confidence that these subcommands can accurately identify true orphans is presently low.

Second, these subcommands store intermediate results on the cluster itself. This can be problematic when cluster administrators are confronting insufficient storage space and want to remove orphans as a means of addressing the issue. The intermediate results could strain the existing cluster storage capacity even further.

For these reasons “orphans find” has been deprecated.

## Orphan List

---

Because “orphans find” has been deprecated, RGW now includes an additional tool - ‘rgw-orphan-list’. When run it will list the available pools and prompt the user to enter the name of the data pool. At that point the tool will, perhaps after an extended period of time, produce a local file containing the RADOS objects from the designated pool that appear to be orphans. The administrator is free to examine this file and decide on a course of action, perhaps removing those RADOS objects from the designated pool.

All intermediate results are stored on the local file system rather than the Ceph cluster. So running the ‘rgw-orphan-list’ tool should have no appreciable impact on the amount of cluster storage consumed.

### WARNING: Experimental Status

The ‘rgw-orphan-list’ tool is new and therefore currently considered experimental. The list of orphans produced should be “sanity checked” before being used for a large delete operation.

### WARNING: Specifying a Data Pool

If a pool other than an RGW data pool is specified, the results of the tool will be erroneous. All RADOS objects found on such a pool will falsely be designated as orphans.

### WARNING: Unindexed Buckets

RGW allows for unindexed buckets, that is buckets that do not maintain an index of their contents. This is not a typical configuration, but it is supported. Because the ‘rgw-orphan-list’ tool uses the bucket indices to determine what RADOS objects should exist, objects in the unindexed buckets will falsely be listed as orphans.

## RADOS List

---

One of the sub-steps in computing a list of orphans is to map each RGW object into its corresponding set of RADOS objects. This is done using a subcommand of ‘radosgw-admin’.

::

```
# radosgw-admin bucket radoslist [-bucket={bucket-name}]
```

The subcommand will produce a list of RADOS objects that support all of the RGW objects. If a bucket is specified then the subcommand will only produce a list of RADOS objects that correspond back the RGW objects in the specified bucket.

## Note: Shared Bucket Markers

Some administrators will be aware of the coding schemes used to name the RADOS objects that correspond to RGW objects, which include a “marker” unique to a given bucket.

RADOS objects that correspond with the contents of one RGW bucket, however, may contain a marker that specifies a different bucket. This behavior is a consequence of the “shallow copy” optimization used by RGW. When larger objects are copied from bucket to bucket, only the “head” objects are actually copied, and the tail objects are shared. Those shared objects will contain the marker of the original bucket.

# OpenID Connect Provider in RGW

An entity describing the OpenID Connect Provider needs to be created in RGW, in order to establish trust between the two.

## REST APIs for Manipulating an OpenID Connect Provider

The following REST APIs can be used for creating and managing an OpenID Connect Provider entity in RGW.

In order to invoke the REST admin APIs, a user with admin caps needs to be created.

```
1. radosgw-admin --uid TESTER --display-name "TestUser" --access_key TESTER --secret test123 user create  
2. radosgw-admin caps add --uid="TESTER" --caps="oidc-provider=*"
```

### CreateOpenIDConnectProvider

Create an OpenID Connect Provider entity in RGW

#### Request Parameters

`ClientIDList.member.N`

Description

List of Client Ids that needs access to S3 resources.

Type

Array of Strings

`ThumbprintList.member.N`

Description

List of OpenID Connect IDP's server certificates' thumbprints. A maximum of 5 thumbprints are allowed.

Type

Array of Strings

`Url`

Description

URL of the IDP.

Type

String

Example::

- POST "<hostname>?Action=CreateOpenIDConnectProvider  
&ThumbprintList.list.1=F7D7B3515DD0D319DD219A43A9EA727AD6065287  
&ClientIDList.list.1=app-profile-jsp  
&Url=<http://localhost:8080/auth/realm/quickstart>

## DeleteOpenIDConnectProvider

Deletes an OpenID Connect Provider entity in RGW

### Request Parameters

`OpenIDConnectProviderArn`

Description

ARN of the IDP which is returned by the Create API.

Type

String

Example::

- POST "<hostname>?Action=DeleteOpenIDConnectProvider  
&OpenIDConnectProviderArn=arn:aws:iam:::oidc-provider/localhost:8080/auth/realm/quickstart

## GetOpenIDConnectProvider

Gets information about an IDP.

### Request Parameters

`OpenIDConnectProviderArn`

Description

ARN of the IDP which is returned by the Create API.

Type

String

Example::

- POST "<hostname>?Action=Action=GetOpenIDConnectProvider&OpenIDConnectProviderArn=arn:aws:iam::oidc-provider/localhost:8080/auth/realmms/quickstart"

## ListOpenIDConnectProviders

Lists information about all IDPs

### Request Parameters

None

Example::

POST "<hostname>?Action=Action=ListOpenIDConnectProviders

# Troubleshooting

## The Gateway Won't Start

If you cannot start the gateway (i.e., there is no existing `pid`), check to see if there is an existing `.asok` file from another user. If an `.asok` file from another user exists and there is no running `pid`, remove the `.asok` file and try to start the process again. This may occur when you start the process as a `root` user and the startup script is trying to start the process as a `www-data` or `apache` user and an existing `.asok` is preventing the script from starting the daemon.

The radosgw init script (`/etc/init.d/radosgw`) also has a verbose argument that can provide some insight as to what could be the issue:

```
1. /etc/init.d/radosgw start -v
```

or

```
1. /etc/init.d radosgw start --verbose
```

## HTTP Request Errors

Examining the access and error logs for the web server itself is probably the first step in identifying what is going on. If there is a 500 error, that usually indicates a problem communicating with the `radosgw` daemon. Ensure the daemon is running, its socket path is configured, and that the web server is looking for it in the proper location.

## Crashed `radosgw` process

If the `radosgw` process dies, you will normally see a 500 error from the web server (apache, nginx, etc.). In that situation, simply restarting radosgw will restore service.

To diagnose the cause of the crash, check the log in `/var/log/ceph` and/or the core file (if one was generated).

## Blocked `radosgw` Requests

If some (or all) radosgw requests appear to be blocked, you can get some insight into the internal state of the `radosgw` daemon via its admin socket. By default, there will be a socket configured to reside in `/var/run/ceph`, and the daemon can be queried with:

```

1. ceph daemon /var/run/ceph/client.rgw help
2.
3. help           list available commands
4. objecter_requests show in-progress osd requests
5. perfcounters_dump dump perfcounters value
6. perfcounters_schema dump perfcounters schema
7. version        get protocol version

```

of particular interest:

```

1. ceph daemon /var/run/ceph/client.rgw objecter_requests
2. ...

```

will dump information about current in-progress requests with the RADOS cluster. This allows one to identify if any requests are blocked by a non-responsive OSD. For example, one might see:

```

1. { "ops": [
2.     { "tid": 1858,
3.         "pg": "2.d2041a48",
4.         "osd": 1,
5.         "last_sent": "2012-03-08 14:56:37.949872",
6.         "attempts": 1,
7.         "object_id": "fatty_25647_object1857",
8.         "object_locator": "@2",
9.         "snapid": "head",
10.        "snap_context": "0=[]",
11.        "mtime": "2012-03-08 14:56:37.949813",
12.        "osd_ops": [
13.            "write 0~4096"]},
14.        { "tid": 1873,
15.            "pg": "2.695e9f8e",
16.            "osd": 1,
17.            "last_sent": "2012-03-08 14:56:37.970615",
18.            "attempts": 1,
19.            "object_id": "fatty_25647_object1872",
20.            "object_locator": "@2",
21.            "snapid": "head",
22.            "snap_context": "0=[]",
23.            "mtime": "2012-03-08 14:56:37.970555",
24.            "osd_ops": [
25.                "write 0~4096"]}],
26.        "linger_ops": [],
27.        "pool_ops": [],
28.        "pool_stat_ops": [],
29.        "statfs_ops": []}

```

In this dump, two requests are in progress. The `last_sent` field is the time the RADOS request was sent. If this is a while ago, it suggests that the OSD is not responding. For example, for request 1858, you could check the OSD status with:

```

1. ceph pg map 2.d2041a48
2.
3. osdmap e9 pg 2.d2041a48 (2.0) -> up [1,0] acting [1,0]

```

This tells us to look at `osd.1`, the primary copy for this PG:

```

1. ceph daemon osd.1 ops
2. { "num_ops": 651,
3.   "ops": [
4.     { "description": "osd_op(client.4124.0:1858 fatty_25647_object1857 [write 0~4096] 2.d2041a48)",
5.      "received_at": "1331247573.344650",
6.      "age": "25.606449",
7.      "flag_point": "waiting for sub ops",
8.      "client_info": { "client": "client.4124",
9.                      "tid": 1858}},
10.    ...

```

The `flag_point` field indicates that the OSD is currently waiting for replicas to respond, in this case `osd.0`.

## Java S3 API Troubleshooting

### Peer Not Authenticated

You may receive an error that looks like this:

```
1. [java] INFO: Unable to execute HTTP request: peer not authenticated
```

The Java SDK for S3 requires a valid certificate from a recognized certificate authority, because it uses HTTPS by default. If you are just testing the Ceph Object Storage services, you can resolve this problem in a few ways:

- Prepend the IP address or hostname with `http://`. For example, change this:

```
1. conn.setEndpoint("myserver");
```

To:

```
1. conn.setEndpoint("http://myserver")
```

- After setting your credentials, add a client configuration and set the protocol to `Protocol.HTTP`.

```

1. AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
2.
3. ClientConfiguration clientConfig = new ClientConfiguration();

```

```

4. clientConfig.setProtocol(Protocol.HTTP);
5.
6. AmazonS3 conn = new AmazonS3Client(credentials, clientConfig);

```

## 405 MethodNotAllowed

If you receive an 405 error, check to see if you have the S3 subdomain set up correctly. You will need to have a wild card setting in your DNS record for subdomain functionality to work properly.

Also, check to ensure that the default site is disabled.

```

[java] Exception in thread "main" Status Code: 405, AWS Service: Amazon S3, AWS Request ID: null, AWS Error
1. Code: MethodNotAllowed, AWS Error Message: null, S3 Extended Request ID: null

```

## Numerous objects in default.rgw.meta pool

Clusters created prior to *jewel* have a metadata archival feature enabled by default, using the `default.rgw.meta` pool. This archive keeps all old versions of user and bucket metadata, resulting in large numbers of objects in the `default.rgw.meta` pool.

## Disabling the Metadata Heap

Users who want to disable this feature going forward should set the `metadata_heap` field to an empty string `""` :

```

1. $ radosgw-admin zone get --rgw-zone=default > zone.json
2. [edit zone.json, setting "metadata_heap": ""]
3. $ radosgw-admin zone set --rgw-zone=default --infile=zone.json
4. $ radosgw-admin period update --commit

```

This will stop new metadata from being written to the `default.rgw.meta` pool, but does not remove any existing objects or pool.

## Cleaning the Metadata Heap Pool

Clusters created prior to *jewel* normally use `default.rgw.meta` only for the metadata archival feature.

However, from *luminous* onwards, radosgw uses [Pool Namespaces](#) within `default.rgw.meta` for an entirely different purpose, that is, to store `user_keys` and other critical metadata.

Users should check zone configuration before proceeding any cleanup procedures:

```
1. $ radosgw-admin zone get --rgw-zone=default | grep default.rgw.meta
```

2. [should not match any strings]

Having confirmed that the pool is not used for any purpose, users may safely delete all objects in the `default.rgw.meta` pool, or optionally, delete the entire pool itself.

# radosgw – rados REST gateway

---

## Synopsis

```
radosgw
```

## Description

**radosgw** is an HTTP REST gateway for the RADOS object store, a part of the Ceph distributed storage system. It is implemented as a FastCGI module using libfcgi, and can be used in conjunction with any FastCGI capable web server.

## Options

```
-c ceph.conf, --conf``=ceph.conf
```

Use **ceph.conf** configuration file instead of the default **/etc/ceph/ceph.conf** to determine monitor addresses during startup.

```
-m monaddress[:port]
```

Connect to specified monitor (instead of looking through **ceph.conf**).

```
-i ID, --id ID
```

Set the ID portion of name for radosgw

```
-n TYPE.ID, --name TYPE.ID
```

Set the rados user name for the gateway (eg. `client.radosgw.gateway`)

```
--cluster NAME
```

Set the cluster name (default: ceph)

```
-d
```

Run in foreground, log to stderr

```
-f
```

Run in foreground, log to usual location

```
--rgw-socket-path``=path
```

Specify a unix domain socket path.

```
--rgw-region``=region
```

The region where radosgw runs

```
--rgw-zone``=zone
```

The zone where radosgw runs

## Configuration

Earlier RADOS Gateway had to be configured with `Apache` and `mod_fastcgi`. Now, `mod_proxy_fcgi` module is used instead of `mod_fastcgi`. `mod_proxy_fcgi` works differently than a traditional FastCGI module. This module requires the service of `mod_proxy` which provides support for the FastCGI protocol. So, to be able to handle FastCGI protocol, both `mod_proxy` and `mod_proxy_fcgi` have to be present in the server. Unlike `mod_fastcgi`, `mod_proxy_fcgi` cannot start the application process. Some platforms have `fcgistarterm` for that purpose. However, external launching of application or process management may be available in the FastCGI application framework in use.

`Apache` can be configured in a way that enables `mod_proxy_fcgi` to be used with localhost tcp or through unix domain socket. `mod_proxy_fcgi` that doesn't support unix domain socket such as the ones in Apache 2.2 and earlier versions of Apache 2.4, needs to be configured for use with localhost tcp. Later versions of Apache like Apache 2.4.9 or later support unix domain socket and as such they allow for the configuration with unix domain socket instead of localhost tcp.

The following steps show the configuration in Ceph's configuration file i.e., `/etc/ceph/ceph.conf` and the gateway configuration file i.e., `/etc/httpd/conf.d/rgw.conf` (RPM-based distros) or `/etc/apache2/conf-available/rgw.conf` (Debian-based distros) with localhost tcp and through unix domain socket:

- For distros with Apache 2.2 and early versions of Apache 2.4 that use localhost TCP and do not support Unix Domain Socket, append the following contents to

`/etc/ceph/ceph.conf` :

```
1. [client.radosgw.gateway]
2. host = {hostname}
3. keyring = /etc/ceph/ceph.client.radosgw.keyring
4. rgw socket path = ""
5. log file = /var/log/ceph/client.radosgw.gateway.log
6. rgw frontends = fastcgi socket_port=9000 socket_host=0.0.0.0
7. rgw print continue = false
```

- Add the following content in the gateway configuration file:

For Debian/Ubuntu add in `/etc/apache2/conf-available/rgw.conf` :

```
1. <VirtualHost *:80>
2. ServerName localhost
3. DocumentRoot /var/www/html
4.
```

```

5. ErrorLog /var/log/apache2/rgw_error.log
6. CustomLog /var/log/apache2/rgw_access.log combined
7.
8. # LogLevel debug
9.
10. RewriteEngine On
11.
12. RewriteRule .* - [E=HTTP_AUTHORIZATION:%{HTTP:Authorization},L]
13.
14. SetEnv proxy-nokeepalive 1
15.
16. ProxyPass / fcgi://localhost:9000/
17.
18. </VirtualHost>

```

For CentOS/RHEL add in `/etc/httpd/conf.d/rgw.conf` :

```

1. <VirtualHost *:80>
2. ServerName localhost
3. DocumentRoot /var/www/html
4.
5. ErrorLog /var/log/httpd/rgw_error.log
6. CustomLog /var/log/httpd/rgw_access.log combined
7.
8. # LogLevel debug
9.
10. RewriteEngine On
11.
12. RewriteRule .* - [E=HTTP_AUTHORIZATION:%{HTTP:Authorization},L]
13.
14. SetEnv proxy-nokeepalive 1
15.
16. ProxyPass / fcgi://localhost:9000/
17.
18. </VirtualHost>

```

3. For distros with Apache 2.4.9 or later that support Unix Domain Socket, append the following configuration to `/etc/ceph/ceph.conf` :

```

1. [client.radosgw.gateway]
2. host = {hostname}
3. keyring = /etc/ceph/ceph.client.radosgw.keyring
4. rgw socket path = /var/run/ceph/ceph.radosgw.gateway.fastcgi.sock
5. log file = /var/log/ceph/client.radosgw.gateway.log
6. rgw print continue = false

```

4. Add the following content in the gateway configuration file:

For CentOS/RHEL add in `/etc/httpd/conf.d/rgw.conf` :

```

1. <VirtualHost *:80>

```

```

2. ServerName localhost
3. DocumentRoot /var/www/html
4.
5. ErrorLog /var/log/httpd/rgw_error.log
6. CustomLog /var/log/httpd/rgw_access.log combined
7.
8. # LogLevel debug
9.
10. RewriteEngine On
11.
12. RewriteRule .* - [E=HTTP_AUTHORIZATION:{HTTP:Authorization},L]
13.
14. SetEnv proxy-nokeepalive 1
15.
16. ProxyPass / unix:///var/run/ceph/ceph.radosgw.gateway.fastcgi.sock|fcgi://localhost:9000/
17.
18. </VirtualHost>

```

Please note, [Apache 2.4.7](#) does not have Unix Domain Socket support in it and as such it has to be configured with localhost tcp. The Unix Domain Socket support is available in [Apache 2.4.9](#) and later versions.

## 5. Generate a key for radosgw to use for authentication with the cluster.

```

1. ceph-authtool -C -n client.radosgw.gateway --gen-key /etc/ceph/keyring.radosgw.gateway
   ceph-authtool -n client.radosgw.gateway --cap mon 'allow rw' --cap osd 'allow rwx'
2. /etc/ceph/keyring.radosgw.gateway

```

## 6. Add the key to the auth entries.

```
1. ceph auth add client.radosgw.gateway --in-file=keyring.radosgw.gateway
```

## 7. Start Apache and radosgw.

Debian/Ubuntu:

```

1. sudo /etc/init.d/apache2 start
2. sudo /etc/init.d/radosgw start

```

CentOS/RHEL:

```

1. sudo apachectl start
2. sudo /etc/init.d/ceph-radosgw start

```

# Usage Logging

**radosgw** maintains an asynchronous usage log. It accumulates statistics about user operations and flushes it periodically. The logs can be accessed and managed through

## radosgw-admin.

The information that is being logged contains total data transfer, total operations, and total successful operations. The data is being accounted in an hourly resolution under the bucket owner, unless the operation was done on the service (e.g., when listing a bucket) in which case it is accounted under the operating user.

Following is an example configuration:

```
1. [client.radosgw.gateway]
2.   rgw enable usage log = true
3.   rgw usage log tick interval = 30
4.   rgw usage log flush threshold = 1024
5.   rgw usage max shards = 32
6.   rgw usage max user shards = 1
```

The total number of shards determines how many total objects hold the usage log information. The per-user number of shards specify how many objects hold usage information for a single user. The tick interval configures the number of seconds between log flushes, and the flush threshold specify how many entries can be kept before resorting to synchronous flush.

## Availability

**radosgw** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[ceph\(8\)](#) [radosgw-admin\(8\)](#)

# radosgw-admin – rados REST gateway user administration utility

---

## Synopsis

---

**radosgw-admin** *command* [ *options* ... ]

## Description

---

**radosgw-admin** is a RADOS gateway user administration utility. It allows creating and modifying users.

## Commands

---

**radosgw-admin** utility uses many commands for administration purpose which are as follows:

### **user create**

Create a new user.

### **user modify**

Modify a user.

### **user info**

Display information of a user, and any potentially available subusers and keys.

### **user rename**

Renames a user.

### **user rm**

Remove a user.

### **user suspend**

Suspend a user.

### **user enable**

Re-enable user after suspension.

### **user check**

Check user info.

### **user stats**

Show user stats as accounted by quota subsystem.

### **user list**

List all users.

### **caps add**

Add user capabilities.

### **caps rm**

Remove user capabilities.

### **subuser create**

Create a new subuser (primarily useful for clients using the Swift API).

### **subuser modify**

Modify a subuser.

### **subuser rm**

Remove a subuser.

### **key create**

Create access key.

### **key rm**

Remove access key.

### **bucket list**

List buckets, or, if bucket specified with -bucket=<bucket>, list its objects. If bucket specified adding -allow-unordered removes ordering requirement, possibly generating results more quickly in buckets with large number of objects.

### **bucket limit check**

Show bucket sharding stats.

### **bucket link**

Link bucket to specified user.

### **bucket unlink**

Unlink bucket from specified user.

### **bucket chown**

Link bucket to specified user and update object ACLs. Use -marker to resume if command gets interrupted.

### **bucket stats**

Returns bucket statistics.

### **bucket rm**

Remove a bucket.

### **bucket check**

Check bucket index.

### **bucket rewrite**

Rewrite all objects in the specified bucket.

### **bucket radoslist**

List the rados objects that contain the data for all objects in the designated bucket, if -bucket=<bucket> is specified, or otherwise all buckets.

### **bucket reshard**

Reshard a bucket.

### **bucket sync disable**

Disable bucket sync.

### **bucket sync enable**

Enable bucket sync.

### **bi get**

Retrieve bucket index object entries.

### **bi put**

Store bucket index object entries.

### **bi list**

List raw bucket index entries.

### **bi purge**

Purge bucket index entries.

**object rm**

Remove an object.

**object stat**

Stat an object for its metadata.

**object unlink**

Unlink object from bucket index.

**object rewrite**

Rewrite the specified object.

**objects expire**

Run expired objects cleanup.

**period rm**

Remove a period.

**period get**

Get the period info.

**period get-current**

Get the current period info.

**period pull**

Pull a period.

**period push**

Push a period.

**period list**

List all periods.

**period update**

Update the staging period.

**period commit**

Commit the staging period.

**quota set**

Set quota params.

**quota enable**

Enable quota.

**quota disable**

Disable quota.

**global quota get**

View global quota parameters.

**global quota set**

Set global quota parameters.

**global quota enable**

Enable a global quota.

**global quota disable**

Disable a global quota.

**realm create**

Create a new realm.

**realm rm**

Remove a realm.

**realm get**

Show the realm info.

**realm get-default**

Get the default realm name.

**realm list**

List all realms.

**realm list-periods**

List all realm periods.

**realm rename**

Rename a realm.

### **realm set**

Set the realm info (requires infile).

### **realm default**

Set the realm as default.

### **realm pull**

Pull a realm and its current period.

### **zonegroup add**

Add a zone to a zonegroup.

### **zonegroup create**

Create a new zone group info.

### **zonegroup default**

Set the default zone group.

### **zonegroup rm**

Remove a zone group info.

### **zonegroup get**

Show the zone group info.

### **zonegroup modify**

Modify an existing zonegroup.

### **zonegroup set**

Set the zone group info (requires infile).

### **zonegroup remove**

Remove a zone from a zonegroup.

### **zonegroup rename**

Rename a zone group.

### **zonegroup list**

List all zone groups set on this cluster.

**zonegroup placement list**

List zonegroup's placement targets.

**zonegroup placement add**

Add a placement target id to a zonegroup.

**zonegroup placement modify**

Modify a placement target of a specific zonegroup.

**zonegroup placement rm**

Remove a placement target from a zonegroup.

**zonegroup placement default**

Set a zonegroup's default placement target.

**zone create**

Create a new zone.

**zone rm**

Remove a zone.

**zone get**

Show zone cluster params.

**zone set**

Set zone cluster params (requires infile).

**zone modify**

Modify an existing zone.

**zone list**

List all zones set on this cluster.

**metadata sync status**

Get metadata sync status.

**metadata sync init**

Init metadata sync.

**metadata sync run**

Run metadata sync.

**data sync status**

Get data sync status of the specified source zone.

**data sync init**

Init data sync for the specified source zone.

**data sync run**

Run data sync for the specified source zone.

**sync error list**

List sync error.

**sync error trim**

Trim sync error.

**zone rename**

Rename a zone.

**zone placement list**

List zone's placement targets.

**zone placement add**

Add a zone placement target.

**zone placement modify**

Modify a zone placement target.

**zone placement rm**

Remove a zone placement target.

**pool add**

Add an existing pool for data placement.

**pool rm**

Remove an existing pool from data placement set.

**pools list**

List placement active set.

## **policy**

Display bucket/object policy.

## **log list**

List log objects.

## **log show**

Dump a log from specific object or (bucket + date + bucket-id). (NOTE: required to specify formatting of date to "YYYY-MM-DD-hh")

## **log rm**

Remove log object.

## **usage show**

Show the usage information (with optional user and date range).

## **usage trim**

Trim usage information (with optional user and date range).

## **gc list**

Dump expired garbage collection objects (specify -include-all to list all entries, including unexpired).

## **gc process**

Manually process garbage.

## **lc list**

List all bucket lifecycle progress.

## **lc process**

Manually process lifecycle.

## **metadata get**

Get metadata info.

## **metadata put**

Put metadata info.

## **metadata rm**

Remove metadata info.

**metadata list**

List metadata info.

**mdlog list**

List metadata log.

**mdlog trim**

Trim metadata log.

**mdlog status**

Read metadata log status.

**bilog list**

List bucket index log.

**bilog trim**

Trim bucket index log (use start-marker, end-marker).

**datalog list**

List data log.

**datalog trim**

Trim data log.

**datalog status**

Read data log status.

**orphans find**

Init and run search for leaked rados objects. DEPRECATED. See the “rgw-orphan-list” tool.

**orphans finish**

Clean up search for leaked rados objects. DEPRECATED. See the “rgw-orphan-list” tool.

**orphans list-jobs**

List the current job-ids for the orphans search. DEPRECATED. See the “rgw-orphan-list” tool.

**role create**

Create a new AWS role for use with STS.

**role rm**

Remove a role.

**role get**

Get a role.

**role list**

List the roles with specified path prefix.

**role modify**

Modify the assume role policy of an existing role.

**role-policy put**

Add/update permission policy to role.

**role-policy list**

List the policies attached to a role.

**role-policy get**

Get the specified inline policy document embedded with the given role.

**role-policy rm**

Remove the policy attached to a role

**reshard add**

Schedule a resharding of a bucket

**reshard list**

List all bucket resharding or scheduled to be resharded

**reshard process**

Process of scheduled reshard jobs

**reshard status**

Resharding status of a bucket

**reshard cancel**

Cancel resharding a bucket

**topic list**

List bucket notifications/pubsub topics

### **topic get**

Get a bucket notifications/pubsub topic

### **topic rm**

Remove a bucket notifications/pubsub topic

### **subscription get**

Get a pubsub subscription definition

### **subscription rm**

Remove a pubsub subscription

### **subscription pull**

Show events in a pubsub subscription

### **subscription ack**

Ack (remove) an events in a pubsub subscription

## Options

---

`-c ceph.conf` ` --conf` `=ceph.conf`

Use `ceph.conf` configuration file instead of the default `/etc/ceph/ceph.conf` to determine monitor addresses during startup.

`-m monaddress[:port]`

Connect to specified monitor (instead of looking through `ceph.conf`).

`--tenant` `=<tenant>`

Name of the tenant.

`--uid` `=uid`

The radosgw user ID.

`--new-uid` `=uid`

ID of the new user. Used with ‘user rename’ command.

`--subuser` `=<name>`

Name of the subuser.

`--access-key` `=<key>`

S3 access key.

--email``=email

The e-mail address of the user.

--secret//--secret-key``=<key>

The secret key.

--gen-access-key

Generate random access key (for S3).

--gen-secret

Generate random secret key.

--key-type``=<type>

key type, options are: swift, s3.

--temp-url-key``[-2]=<key>

Temporary url key.

--max-buckets

max number of buckets for a user (0 for no limit, negative value to disable bucket creation). Default is 1000.

--access``=<access>

Set the access permissions for the sub-user. Available access permissions are read, write, readwrite and full.

--display-name``=<name>

The display name of the user.

--admin

Set the admin flag on the user.

--system

Set the system flag on the user.

--bucket``=[tenant-id/]bucket

Specify the bucket name. If tenant-id is not specified, the tenant-id of the user (-uid) is used.

--pool``=<pool>

Specify the pool name. Also used with orphans find as data pool to scan for leaked rados objects.

`--object``=object`

Specify the object name.

`--date``=yyyy-mm-dd`

The date in the format yyyy-mm-dd.

`--start-date``=yyyy-mm-dd`

The start date in the format yyyy-mm-dd.

`--end-date``=yyyy-mm-dd`

The end date in the format yyyy-mm-dd.

`--bucket-id``=<bucket-id>`

Specify the bucket id.

`--bucket-new-name``=[tenant-id/]<bucket>`

- Optional for bucket link; use to rename a bucket.

While tenant-id/ can be specified, this is never necessary for normal operation.

`--shard-id``=<shard-id>`

Optional for mdlog list, bi list, data sync status. Required for `mdlog trim`.

`--max-entries``=<entries>`

Optional for listing operations to specify the max entires

`--purge-data`

When specified, user removal will also purge all the user data.

`--purge-keys`

When specified, subuser removal will also purge all the subuser keys.

`--purge-objects`

When specified, the bucket removal will also purge all objects in it.

`--metadata-key``=<key>`

Key to retrieve metadata from with `metadata get`.

`--remote``=<remote>`

Zone or zonegroup id of remote gateway.

`--period``=<id>`

Period id.

`--url``=<url>`

url for pushing/pulling period or realm.

--epoch` `=<number>

Period epoch.

--commit

Commit the period during 'period update'.

--staging

Get the staging period info.

--master

Set as master.

--master-zone` `=<id>

Master zone id.

--rgw-realm` `=<name>

The realm name.

--realm-id` `=<id>

The realm id.

--realm-new-name` `=<name>

New name of realm.

--rgw-zonegroup` `=<name>

The zonegroup name.

--zonegroup-id` `=<id>

The zonegroup id.

--zonegroup-new-name` `=<name>

The new name of the zonegroup.

--rgw-zone` `=<zone>

Zone in which radosgw is running.

--zone-id` `=<id>

The zone id.

--zone-new-name` `=<name>

The new name of the zone.

--source-zone

The source zone for data sync.

--default

Set the entity (realm, zonegroup, zone) as default.

--read-only

Set the zone as read-only when adding to the zonegroup.

--placement-id

Placement id for the zonegroup placement commands.

--tags``=<list>

The list of tags for zonegroup placement add and modify commands.

--tags-add``=<list>

The list of tags to add for zonegroup placement modify command.

--tags-rm``=<list>

The list of tags to remove for zonegroup placement modify command.

--endpoints``=<list>

The zone endpoints.

--index-pool``=<pool>

The placement target index pool.

--data-pool``=<pool>

The placement target data pool.

--data-extra-pool``=<pool>

The placement target data extra (non-ec) pool.

--placement-index-type``=<type>

The placement target index type (normal, indexless, or #id).

--tier-type``=<type>

The zone tier type.

--tier-config``=<k>=<v>[, ...]

Set zone tier config keys, values.

--tier-config-rm``=<k>[, ...]

Unset zone tier config keys.

--sync-from-all``[=false]

Set/reset whether zone syncs from all zonegroup peers.

```
--sync-from``=[zone-name][,...]
```

Set the list of zones to sync from.

```
--sync-from-rm``=[zone-name][,...]
```

Remove the zones from list of zones to sync from.

```
--bucket-index-max-shards
```

Override a zone's or zonegroup's default number of bucket index shards. This option is accepted by the 'zone create', 'zone modify', 'zonegroup add', and 'zonegroup modify' commands, and applies to buckets that are created after the zone/zonegroup changes take effect.

```
--fix
```

Besides checking bucket index, will also fix it.

```
--check-objects
```

bucket check: Rebuilds bucket index according to actual objects state.

```
--format``=<format>
```

Specify output format for certain operations. Supported formats: xml, json.

```
--sync-stats
```

Option for 'user stats' command. When specified, it will update user stats with the current stats reported by user's buckets indexes.

```
--show-log-entries``=<flag>
```

Enable/disable dump of log entries on log show.

```
--show-log-sum``=<flag>
```

Enable/disable dump of log summation on log show.

```
--skip-zero-entries
```

Log show only dumps entries that don't have zero value in one of the numeric field.

```
--infile
```

Specify a file to read in when setting data.

```
--categories``=<list>
```

Comma separated list of categories, used in usage show.

```
--caps``=<caps>
```

List of caps (e.g., "usage=read, write; user=read").

`--compression``=<compression-algorithm>`

Placement target compression algorithm (lz4|snappy|zlib|zstd)

`--yes-i-really-mean-it`

Required for certain operations.

`--min-rewrite-size`

Specify the min object size for bucket rewrite (default 4M).

`--max-rewrite-size`

Specify the max object size for bucket rewrite (default ULONG\_MAX).

`--min-rewrite-stripe-size`

Specify the min stripe size for object rewrite (default 0). If the value is set to 0, then the specified object will always be rewritten for restriping.

`--warnings-only`

When specified with bucket limit check, list only buckets nearing or over the current max objects per shard value.

`--bypass-gc`

When specified with bucket deletion, triggers object deletions by not involving GC.

`--inconsistent-index`

When specified with bucket deletion and bypass-gc set to true, ignores bucket index consistency.

`--max-concurrent-ios`

Maximum concurrent ios for bucket operations. Affects operations that scan the bucket index, e.g., listing, deletion, and all scan/search operations such as finding orphans or checking the bucket index. Default is 32.

## Quota Options

`--max-objects`

Specify max objects (negative value to disable).

`--max-size`

Specify max size (in B/K/M/G/T, negative value to disable).

`--quota-scope`

The scope of quota (bucket, user).

## Orphans Search Options

---

--num-shards

Number of shards to use for keeping the temporary scan info

--orphan-stale-secs

Number of seconds to wait before declaring an object to be an orphan. Default is 86400 (24 hours).

--job-id

Set the job id (for orphans find)

## Orphans list-jobs options

---

--extra-info

Provide extra info in the job list.

## Role Options

---

--role-name

The name of the role to create.

--path

The path to the role.

--assume-role-policy-doc

The trust relationship policy document that grants an entity permission to assume the role.

--policy-name

The name of the policy document.

--policy-doc

The permission policy document.

--path-prefix

The path prefix for filtering the roles.

## Bucket Notifications/PubSub Options

---

--topic

The bucket notifications/pubsub topic name.

```
--subscription
```

The pubsub subscription name.

```
--event-id
```

The event id in a pubsub subscription.

## Examples

Generate a new user:

```
1. $ radosgw-admin user create --display-name="johnny rotten" --uid=johnny
2. { "user_id": "johnny",
3.   "rados_uid": 0,
4.   "display_name": "johnny rotten",
5.   "email": "",
6.   "suspended": 0,
7.   "subusers": [],
8.   "keys": [
9.     { "user": "johnny",
10.       "access_key": "TCICW53D9BQ2VGC46I44",
11.       "secret_key": "tfm9aHMI8X76L3UdgE+ZQaJag1vJQmE6HDb5Lbrz"}],
12.   "swift_keys": []}
```

Remove a user:

```
1. $ radosgw-admin user rm --uid=johnny
```

Rename a user:

```
1. $ radosgw-admin user rename --uid=johny --new-uid=joe
```

Remove a user and all associated buckets with their contents:

```
1. $ radosgw-admin user rm --uid=johnny --purge-data
```

Remove a bucket:

```
1. $ radosgw-admin bucket rm --bucket=foo
```

Link bucket to specified user:

```
1. $ radosgw-admin bucket link --bucket=foo --bucket_id=<bucket id> --uid=johnny
```

Unlink bucket from specified user:

```
1. $ radosgw-admin bucket unlink --bucket=foo --uid=johnny
```

Rename a bucket:

```
1. $ radosgw-admin bucket link --bucket=foo --bucket-new-name=bar --uid=johnny
```

Move a bucket from the old global tenant space to a specified tenant:

```
1. $ radosgw-admin bucket link --bucket=/foo --uid=12345678$12345678'
```

Link bucket to specified user and change object ACLs:

```
1. $ radosgw-admin bucket chown --bucket=/foo --uid=12345678$12345678'
```

Show the logs of a bucket from April 1st, 2012:

```
1. $ radosgw-admin log show --bucket=foo --date=2012-04-01 --bucket-id=default.14193.1
```

Show usage information for user from March 1st to (but not including) April 1st, 2012:

```
1. $ radosgw-admin usage show --uid=johnny \
2.           --start-date=2012-03-01 --end-date=2012-04-01
```

Show only summary of usage information for all users:

```
1. $ radosgw-admin usage show --show-log-entries=false
```

Trim usage information for user until March 1st, 2012:

```
1. $ radosgw-admin usage trim --uid=johnny --end-date=2012-04-01
```

## Availability

**radosgw-admin** is part of Ceph, a massively scalable, open-source, distributed storage system. Please refer to the Ceph documentation at <http://ceph.com/docs> for more information.

## See also

[ceph\(8\)](#) [radosgw\(8\)](#)

# QAT Acceleration for Encryption and Compression

Intel QAT (QuickAssist Technology) can provide extended accelerated encryption and compression services by offloading the actual encryption and compression request(s) to the hardware QuickAssist accelerators, which are more efficient in terms of cost and power than general purpose CPUs for those specific compute-intensive workloads.

See [QAT Support for Compression](#) and [QAT based Encryption for RGW](#).

## QAT in the Software Stack

Application developers can access QuickAssist features through the QAT API. The QAT API is the top-level API for QuickAssist technology, and enables easy interfacing between the customer application and the QuickAssist acceleration driver.

The QAT API accesses the QuickAssist driver, which in turn drives the QuickAssist Accelerator hardware. The QuickAssist driver is responsible for exposing the acceleration services to the application software.

A user can write directly to the QAT API, or the use of QAT can be done via frameworks that have been enabled by others including Intel (for example, zlib\*, OpenSSL\*, libcrypto\*, and the Linux\* Kernel Crypto Framework).

## QAT Environment Setup

1. QuickAssist Accelerator hardware is necessary to make use of accelerated encryption and compression services. And QAT driver in kernel space have to be loaded to drive the hardware.

The driver package can be downloaded from [Intel Quickassist Technology](#).

1. The implementation for QAT based encryption is directly base on QAT API which is included the driver package. But QAT support for compression depends on QATzip project, which is a user space library which builds on top of the QAT API. Currently, QATzip speeds up gzip compression and decompression at the time of writing.

See [QATzip](#).

## Implementation

1. QAT based Encryption for RGW

[OpenSSL support for RGW encryption](#) has been merged into Ceph, and Intel also provides one [QAT Engine](#) for OpenSSL. So, theoretically speaking, QAT based encryption in Ceph can be directly supported through OpenSSL+QAT Engine.

But the QAT Engine for OpenSSL currently supports chained operations only, and so Ceph will not be able to utilize QAT hardware feature for crypto operations based on OpenSSL crypto plugin. As a result, one QAT plugin based on native QAT API is added into crypto framework.

### 1. QAT Support for Compression

As mentioned above, QAT support for compression is based on QATzip library in user space, which is designed to take full advantage of the performance provided by QuickAssist Technology. Unlike QAT based encryption, QAT based compression is supported through a tool class for QAT acceleration rather than a compressor plugin. The common tool class can transparently accelerate the existing compression types, but only zlib compressor can be supported at the time of writing. So user is allowed to use it to speed up zlib compressor as long as the QAT hardware is available and QAT is capable to handle it.

## Configuration

### 1. QAT based Encryption for RGW

Edit the Ceph configuration file to make use of QAT based crypto plugin:

```
1. plugin crypto accelerator = crypto_qat
```

### 1. QAT Support for Compression

One CMake option have to be used to trigger QAT based compression:

```
1. -DWITH_QATZIP=ON
```

Edit the Ceph configuration file to enable QAT support for compression:

```
1. qat compressor enabled=true
```

# Ceph s3 select

## Contents

- Ceph s3 select
  - Overview
  - Basic workflow
    - Basic functionalities
    - Error Handling
  - Features Support
  - s3-select function interfaces
    - Timestamp functions
    - Aggregation functions
    - String functions
    - Alias
  - Sending Query to RGW
    - Syntax
  - CSV parsing behavior
  - BOTO3

## Overview

The purpose of the **s3 select** engine is to create an efficient pipe between user client and storage nodes (the engine should be close as possible to storage).

It enables selection of a restricted subset of (structured) data stored in an S3 object using an SQL-like syntax.

It also enables for higher level analytic-applications (such as SPARK-SQL) , using that feature to improve their latency and throughput.

For example, a s3-object of several GB (CSV file), a user needs to extract a single column which filtered by another column.

As the following query:

```
select customer-id from s3object where age>30 and age<65;
```

Currently the whole s3-object must retrieve from OSD via RGW before filtering and extracting data.

By “pushing down” the query into OSD , it’s possible to save a lot of network and CPU(serialization / deserialization).

The bigger the object, and the more accurate the query, the better the performance.

## Basic workflow

S3-select query is sent to RGW via [AWS-CLI](#)

It passes the authentication and permission process as an incoming message (POST).

`RGWSelectObj_ObjStore_S3::send_response_data` is the “entry point”, it handles each fetched chunk according to input object-key.

`send_response_data` is first handling the input query, it extracts the query and other CLI parameters.

Per each new fetched chunk (~4m), RGW executes s3-select query on it.

The current implementation supports CSV objects and since chunks are randomly “cutting” the CSV rows in the middle, those broken-lines (first or last per chunk) are skipped while processing the query.

Those “broken” lines are stored and later merged with the next broken-line (belong to the next chunk), and finally processed.

Per each processed chunk an output message is formatted according to [AWS specification](#) and sent back to the client.

RGW supports the following response:   `{:event-type,records} {:content-type,application/octet-stream} {:message-type,event}` .

For aggregation queries the last chunk should be identified as the end of input, following that the s3-select-engine initiates end-of-process and produces an aggregate result.

## Basic functionalities

S3select has a definite set of functionalities that should be implemented (if we wish to stay compliant with AWS), currently only a portion of it is implemented.

The implemented software architecture supports basic arithmetic expressions, logical and compare expressions, including nested function calls and casting operators, that alone enables the user reasonable flexibility.

review the below [feature-table](#).

## Error Handling

Any error occurs while the input query processing, i.e. parsing phase or execution phase, is returned to client as response error message.

Fatal severity (attached to the exception) will end query execution immediately, other error severity are counted, upon reaching 100, it ends query execution with an error message.

## Features Support

Currently only part of [AWS select command](#) is implemented, table bellow describes what is currently supported.

The following table describes the current implementation for s3-select functionalities:

Feature	Detailed	Example
Arithmetic operators	<code>^ / + - ( )</code>	<code>select (int(_1)+int(_2))int(_9) from stdin;</code> <code>select ((1+2)3.14) ^ 2 from stdin;</code>
Compare operators	<code>&gt; &lt; &gt;= &lt;= == !=</code>	<code>select _1,_2 from stdin where (int(1)+int(_3))&gt;int(_4)</code>
logical operator	AND OR	<code>select count() from stdin where int(1)&gt;123 and int(2)&lt;345</code>
casting operator	<code>int(expression)</code>	<code>select int(_1),int( 1.2 + 3.4) from stdin;</code>
	<code>float(expression)</code>	<code>select float(1.2) from stdin;</code>
	<code>timestamp(...)</code>	<code>select timestamp("1999:10:10-12:23:44") from stdin;</code>
Aggregation Function	sum	<code>select sum(int(_1)) from stdin;</code>
Aggregation Function	min	<code>select min( int(_1) int(_5) ) from stdin;</code>
Aggregation Function	max	<code>select max(float(_1)),min(int(_5)) from stdin;</code>
Aggregation Function	count	<code>select count() from stdin where (int(1)+int(_3))&gt;int(_4)</code>
Timestamp Functions	extract	<code>select count(*) from stdin where extract("year",timestamp(_1)) &gt; 1950 and extract("year",timestamp(_1)) &lt; 1960;</code>
Timestamp Functions	dateadd	<code>select count(0) from stdin where datediff("year",timestamp(_1),dateadd("day",366,timestamp(_1))) == 1;</code>
Timestamp Functions	datediff	<code>select count(0) from stdin where datediff("month",timestamp(_1),timestamp(_2))) == 1;</code>
Timestamp Functions	utcnow	<code>select count(0) from stdin where datediff("hours",utcnow(),dateadd("day",1,utcnow())) == 1;</code>
String Functions	substr	<code>select count(0) from stdin where int(substr(_1,1,4))&lt;1960;</code>
alias support		<code>select int(_1) as a1, int(_2) as a2 , (a1+a2) as a3 from stdin where a3&gt;100 and a3&lt;300;</code>

# s3-select function interfaces

## Timestamp functions

The `timestamp functionalities` is partially implemented.

the casting operator( `timestamp( string )` ), converts string to timestamp basic type.

Currently it can convert the following pattern `yyyy:mm:dd hh:mi:dd`

`extract( date-part , timestamp)` : function return integer according to date-part extract from input timestamp.

supported date-part : year,month,week,day.

`dateadd(date-part , integer,timestamp)` : function return timestamp, a calculation results of input timestamp and date-part.

supported date-part : year,month,day.

`datediff(date-part,timestamp,timestamp)` : function return an integer, a calculated result for difference between 2 timestamps according to date-part.

supported date-part : year,month,day,hours.

`utcnow()` : return timestamp of current time.

## Aggregation functions

`count()` : return integer according to number of rows matching condition(if such exist).

`sum(expression)` : return a summary of expression per all rows matching condition(if such exist).

`max(expression)` : return the maximal result for all expressions matching condition(if such exist).

`min(expression)` : return the minimal result for all expressions matching condition(if such exist).

## String functions

`substr(string,from,to)` : return a string extract from input string according to from,to inputs.

## Alias

Alias programming-construct is an essential part of s3-select language, it enables much better programming especially with objects containing many columns or in the case of complex queries.

Upon parsing the statement containing alias construct, it replaces alias with reference to correct projection column, on query execution time the reference is evaluated as any other expression.

There is a risk that self(or cyclic) reference may occur causing stack-overflow(endless-loop), for that concern upon evaluating an alias, it is validated for cyclic reference.

Alias also maintains result-cache, meaning upon using the same alias more than once, it's not evaluating the same

expression again(it will return the same result), instead it uses the result from cache.

Of Course, per each new row the cache is invalidated.

## Sending Query to RGW

Any http-client can send s3-select request to RGW, it must be compliant with [AWS Request syntax](#).

Sending s3-select request to RGW using AWS cli, should follow [AWS command reference](#).

below is an example for it.

```
1. aws --endpoint-url http://localhost:8000 s3api select-object-content
2. --bucket {BUCKET-NAME}
3. --expression-type 'SQL'
4. --input-serialization
  '{"CSV": {"FieldDelimiter": "," , "QuoteCharacter": "\"", "RecordDelimiter" : "\n" , "QuoteEscapeCharacter"
5. : "\\\" , "FileHeaderInfo": "USE" }, "CompressionType": "NONE"}'
6. --output-serialization '{"CSV": {}}'
7. --key {OBJECT-NAME}
8. --expression "select count(0) from stdin where int(_1)<10;" output.csv
```

## Syntax

**Input serialization** (Implemented), it let the user define the CSV definitions; the default values are {\n} for row-delimiter {}, for field delimiter, {"} for quote, {\} for escape characters.

it handle the **csv-header-info**, the first row in input object containing the schema.

**Output serialization** is currently not implemented, the same for **compression-type**.

s3-select engine contain a CSV parser, which parse s3-objects as follows.

- each row ends with row-delimiter.
- field-separator separates between adjacent columns, successive field separator define NULL column.
- quote-character overrides field separator, meaning , field separator become as any character between quotes.
- escape character disables any special characters, except for row delimiter.

Below are examples for CSV parsing rules.

## CSV parsing behavior

Feature	Description	input ==> tokens
NULL	successive field delimiter	,,1,,2, ==> {null}{null}{1}{null}{2}{null}
QUOTE	quote character overrides field delimiter	11,22,"a,b,c,d",last ==> {11}{22}{"a,b,c,d"}{last}

Escape	escape char overrides meta-character. escape removed	11,22,str=\\"abcd\\\",str2=\\"123\\\",last ==> {11}{22}{str="abcd",str2="123"}{last}
row delimiter	no close quote, row delimiter is closing line	11,22,a="str,44,55,66 ==> {11}{22}{a="str,44,55,66}"
CSV header info	FileHeaderInfo tag	" <b>USE</b> " value means each token on first line is column-name, " <b>IGNORE</b> " value means to skip the first line

## BOT03

using BOT03 is "natural" and easy due to AWS-cli support.

```

def
1. run_s3select(bucket, key, query, column_delim=",", row_delim="\n", quot_char='"', esc_char='\\', csv_header_info="NONE")
2.     s3 = boto3.client('s3',
3.         endpoint_url=endpoint,
4.         aws_access_key_id=access_key,
5.         region_name=region_name,
6.         aws_secret_access_key=secret_key)
7.
8.
9.
10.    r = s3.select_object_content(
11.        Bucket=bucket,
12.        Key=key,
13.        ExpressionType='SQL',
14.        InputSerialization = {"CSV": {"RecordDelimiter": row_delim, "FieldDelimiter": column_delim, "QuoteEscapeCharacter": esc_char, "QuoteCharacter": quot_char, "FileHeaderInfo": csv_header_info}, "CompressionType": "NONE"}, OutputSerialization = {"CSV": {}},
15.        Expression=query, )
16.
17.
18.    result = ""
19.    for event in r['Payload']:
20.        if 'Records' in event:
21.            records = event['Records'][ 'Payload'].decode('utf-8')
22.            result += records
23.
24.    return result
25.
26.
27.
28.
29. run_s3select(
30.     "my_bucket",
31.     "my_csv_object",
32.     "select int(_1) as a1, int(_2) as a2 , (a1+a2) as a3 from stdin where a3>100 and a3<300;"
```



# Lua Scripting

New in version Pacific.

## Contents

- [Lua Scripting](#)
  - [Script Management via CLI](#)
  - [Context Free Functions](#)
    - [Debug Log](#)
  - [Request Fields](#)
  - [Request Functions](#)
    - [Operations Log](#)
  - [Lua Code Samples](#)

This feature allows users to upload Lua scripts to different context in the radosgw. The two supported context are “preRequest” that will execute a script before the operation was taken, and “postRequest” that will execute after each operation is taken. Script may be uploaded to address requests for users of a specific tenant. The script can access fields in the request and modify some fields. All Lua language features can be used in the script.

## Script Management via CLI

To upload a script:

```
1. # radosgw-admin script put --infile={lua-file} --context={preRequest|postRequest} [--tenant={tenant-name}]
```

To print the content of the script to standard output:

```
1. # radosgw-admin script get --context={preRequest|postRequest} [--tenant={tenant-name}]
```

To remove the script:

```
1. # radosgw-admin script rm --context={preRequest|postRequest} [--tenant={tenant-name}]
```

## Context Free Functions

## Debug Log

The `RGWDebugLog()` function accepts a string and prints it to the debug log with priority 20. Each log message is prefixed `Lua INFO:`. This function has no return value.

## Request Fields

### Warning

This feature is experimental. Fields may be removed or renamed in the future.

### Note

- Although Lua is a case-sensitive language, field names provided by the radosgw are case-insensitive. Function names remain case-sensitive.
- Fields marked “optional” can have a nil value.
- Fields marked as “iterable” can be used by the pairs() function and with the # length operator.
- All table fields can be used with the bracket operator `[ ]`.
- `time` fields are strings with the following format: `%Y-%m-%d %H:%M:%S`.

Field	Type	Description	Iterable	W
<code>Request.RGWOp</code>	string	radosgw operation	no	n
<code>Request.DecodedURI</code>	string	decoded URI	no	n
<code>Request.ContentLength</code>	integer	size of the request	no	n
<code>Request.GenericAttributes</code>	table	string to string generic attributes map	yes	n
<code>Request.Response</code>	table	response to the request	no	n
<code>Request.Response.HTTPStatusCode</code>	integer	HTTP status code	no	y
<code>Request.Response.HTTPStatus</code>	string	HTTP status text	no	y
<code>Request.Response.RGWCode</code>	integer	radosgw error code	no	y
<code>Request.Response.Message</code>	string	response message	no	y

Request.SwiftAccountName	string	swift account name	no	n
Request.Bucket	table	info on the bucket	no	n
Request.Bucket.Tenant	string	tenant of the bucket	no	n
Request.Bucket.Name	string	bucket name	no	n
Request.Bucket.Marker	string	bucket marker (initial id)	no	n
Request.Bucket.Id	string	bucket id	no	n
Request.Bucket.Count	integer	number of objects in the bucket	no	n
Request.Bucket.Size	integer	total size of objects in the bucket	no	n
Request.Bucket.ZoneGroupId	string	zone group of the bucket	no	n
Request.Bucket.CreationTime	time	creation time of the bucket	no	n
Request.Bucket.MTime	time	modification time of the bucket	no	n
Request.Bucket.Quota	table	bucket quota	no	n
Request.Bucket.Quota.MaxValue	integer	bucket quota max size	no	n
Request.Bucket.Quota.MaxObjects	integer	bucket quota max number of objects	no	n
Request.Bucket.Quota.Enabled	boolean	bucket quota is enabled	no	n
Request.Bucket.Quota.Rounded	boolean	bucket quota is rounded to 4K	no	n
Request.Bucket.PlacementRule	table	bucket placement rule	no	n
Request.Bucket.PlacementRule.Name	string	bucket placement rule name	no	n
		bucket placement		

			rule storage class		
Request.Bucket.User	table	bucket owner	no	n	
Request.Bucket.User.Tenant	string	bucket owner tenant	no	n	
Request.Bucket.User.Id	string	bucket owner id	no	n	
Request.Object	table	info on the object	no	n	
Request.Object.Name	string	object name	no	n	
Request.Object.Instance	string	object version	no	n	
Request.Object.Id	string	object id	no	n	
Request.Object.Size	integer	object size	no	n	
Request.Object.MTime	time	object mtime	no	n	
Request.CopyFrom	table	information on copy operation	no	n	
Request.CopyFrom.Tenant	string	tenant of the object copied from	no	n	
Request.CopyFrom.Bucket	string	bucket of the object copied from	no	n	
Request.CopyFrom.Object	table	object copied from. See: Request.Object	no	n	
Request.ObjectOwner	table	object owner	no	n	
Request.ObjectOwner.DisplayName	string	object owner display name	no	n	
Request.ObjectOwner.User	table	object user. See: Request.Bucket.User	no	n	
Request.ZoneGroup.Name	string	name of zone group	no	n	
Request.ZoneGroup.Endpoint	string	endpoint of zone group	no	n	
Request.UserAcl	table	user ACL	no	n	
Request.UserAcl.Owner	table	user ACL owner. See:	no	n	

<code>Request.UserAcl.Owner</code>	table	See: <code>Request.ObjectOwner</code>	no	n
<code>Request.UserAcl.Grants</code>	table	user ACL map of string to grant note: grants without an Id are not presented when iterated and only one of them can be accessed via brackets	yes	n
<code>Request.UserAcl.Grants["&lt;name&gt;"]</code>	table	user ACL grant	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].Type</code>	integer	user ACL grant type	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].User</code>	table	user ACL grant user	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].User.Tenant</code>	table	user ACL grant user tenant	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].User.Id</code>	table	user ACL grant user id	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].GroupType</code>	integer	user ACL grant group type	no	n
<code>Request.UserAcl.Grants["&lt;name&gt;"].Referer</code>	string	user ACL grant referer	no	n
<code>Request.BucketAcl</code>	table	bucket ACL . See: <code>Request.UserAcl</code>	no	n
<code>Request.ObjectAcl</code>	table	object ACL . See: <code>Request.UserAcl</code>	no	n
<code>Request.Environment</code>	table	string to string environment map	yes	n
<code>Request.Policy</code>	table	policy	no	n
<code>Request.Policy.Text</code>	string	policy text	no	n
<code>Request.Policy.Id</code>	string	policy Id	no	n
<code>Request.Policy.Statements</code>	table	list of string statements	yes	n
<code>Request.UserPolicies</code>	table	list of user policies	yes	n

<code>Request.UserPolicies[&lt;index&gt;]</code>	table	user policy. See: <code>Request.Policy</code>	no	n
<code>Request.RGWID</code>	string	radosaw host id: <code>&lt;host&gt;-&lt;zone&gt;-&lt;zonestring&gt;</code>	no	n
<code>Request.HTTP</code>	table	HTTP header	no	n
<code>Request.HTTP.Parameters</code>	table	string to string parameter map	yes	n
<code>Request.HTTP.Resources</code>	table	string to string resource map	yes	n
<code>Request.HTTP.Metadata</code>	table	string to string metadata map	yes	y
<code>Request.HTTP.Host</code>	string	host name	no	n
<code>Request.HTTP.Method</code>	string	HTTP method	no	n
<code>Request.HTTP.URI</code>	string	URI	no	n
<code>Request.HTTP.QueryString</code>	string	HTTP query string	no	n
<code>Request.HTTP.Domain</code>	string	domain name	no	n
<code>Request.Time</code>	time	request time	no	n
<code>Request.Dialect</code>	string	“S3” or “Swift”	no	n
<code>Request.Id</code>	string	request Id	no	n
<code>Request.TransactionId</code>	string	transaction Id	no	n
<code>Request.Tags</code>	table	object tags map	yes	n

## Request Functions

### Operations Log

The `Request.Log()` function prints the requests into the operations log. This function has no parameters. It returns 0 for success and an error code if it fails.

### Lua Code Samples

- Print information on source and destination objects in case of copy:

```

1. function print_object(object)
2.   RGWDebugLog(" Name: " .. object.Name)
3.   RGWDebugLog(" Instance: " .. object.Instance)
4.   RGWDebugLog(" Id: " .. object.Id)
5.   RGWDebugLog(" Size: " .. object.Size)
6.   RGWDebugLog(" MTime: " .. object.MTime)
7. end
8.
9. if Request.CopyFrom and Request.Object and Request.CopyFrom.Object then
10.   RGWDebugLog("copy from object:")
11.   print_object(Request.CopyFrom.Object)
12.   RGWDebugLog("to object:")
13.   print_object(Request.Object)
14. end

```

- Print ACLs via a “generic function”:

```

1. function print_owner(owner)
2.   RGWDebugLog("Owner:")
3.   RGWDebugLog(" Dispaly Name: " .. owner.DisplayName)
4.   RGWDebugLog(" Id: " .. owner.User.Id)
5.   RGWDebugLog(" Tenant: " .. owner.User.Tenant)
6. end
7.
8. function print_acl(acl_type)
9.   index = acl_type .. "ACL"
10.  acl = Request[index]
11.  if acl then
12.    RGWDebugLog(acl_type .. "ACL Owner")
13.    print_owner(acl.Owner)
14.    RGWDebugLog(" there are " .. #acl.Grants .. " grant for owner")
15.    for k,v in pairs(acl.Grants) do
16.      RGWDebugLog("   Grant Key: " .. k)
17.      RGWDebugLog("   Grant Type: " .. v.Type)
18.      RGWDebugLog("   Grant Group Type: " .. v.GroupType)
19.      RGWDebugLog("   Grant Referer: " .. v.Referer)
20.      RGWDebugLog("   Grant User Tenant: " .. v.User.Tenant)
21.      RGWDebugLog("   Grant User Id: " .. v.User.Id)
22.    end
23.  else
24.    RGWDebugLog("no " .. acl_type .. " ACL in request: " .. Request.Id)
25.  end
26. end
27.
28. print_acl("User")
29. print_acl("Bucket")
30. print_acl("Object")

```

- Use of operations log only in case of errors:

```

1. if Request.Response.HTTPStatusCode ~= 200 then
2.   RGWDebugLog("request is bad, use ops log")
3.   rc = Request.Log()
4.   RGWDebugLog("ops log return code: " .. rc)
5. end

```

- Set values into the error message:

```

1. if Request.Response.HTTPStatusCode == 500 then
2.   Request.Response.Message = "<Message> something bad happened :-( </Message>"
3. end

```

- Add metadata to objects that was not originally sent by the client:

In the preRequest context we should add:

```

1. if Request.RGWOp == 'put_obj' then
2.   Request.HTTP.Metadata["x-amz-meta-mydata"] = "my value"
3. end

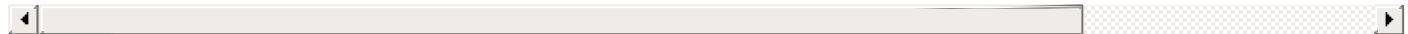
```

In the postRequest context we look at the metadata:

```

1. RGWDebugLog("number of metadata entries is: " .. #Request.HTTP.Metadata)
2. for k, v in pairs(Request.HTTP.Metadata) do
3.   RGWDebugLog("key=" .. k .. ", " .. "value=" .. v)
4. end

```



# Ceph Manager Daemon

The [Ceph Manager](#) daemon (`ceph-mgr`) runs alongside monitor daemons, to provide additional monitoring and interfaces to external monitoring and management systems.

Since the 12.x (*luminous*) Ceph release, the `ceph-mgr` daemon is required for normal operations. The `ceph-mgr` daemon is an optional component in the 11.x (*kraken*) Ceph release.

By default, the manager daemon requires no additional configuration, beyond ensuring it is running. If there is no `mgr` daemon running, you will see a health warning to that effect, and some of the other information in the output of `ceph status` will be missing or stale until a `mgr` is started.

Use your normal deployment tools, such as `ceph-ansible` or `cephadm`, to set up `ceph-mgr` daemons on each of your `mon` nodes. It is not mandatory to place `mgr` daemons on the same nodes as `mons`, but it is almost always sensible.

- [Installation and Configuration](#)
- [Writing modules](#)
- [Writing orchestrator plugins](#)
- [Dashboard module](#)
- [Ceph RESTful API](#)
- [Alerts module](#)
- [DiskPrediction module](#)
- [Local pool module](#)
- [RESTful module](#)
- [Zabbix module](#)
- [Prometheus module](#)
- [Influx module](#)
- [Hello module](#)
- [Telegraf module](#)
- [Telemetry module](#)
- [Iostat module](#)
- [Crash module](#)
- [Insights module](#)
- [Orchestrator module](#)
- [Rook module](#)
- [MDS Autoscaler module](#)

# ceph-mgr administrator's guide

## Manual setup

Usually, you would set up a ceph-mgr daemon using a tool such as ceph-ansible. These instructions describe how to set up a ceph-mgr daemon manually.

First, create an authentication key for your daemon:

```
1. ceph auth get-or-create mgr.$name mon 'allow profile mgr' osd 'allow *' mds 'allow *'
```

Place that key into `mgr data` path, which for a cluster “ceph” and mgr \$name “foo” would be `/var/lib/ceph/mgr/ceph-foo`.

Start the ceph-mgr daemon:

```
1. ceph-mgr -i $name
```

Check that the mgr has come up by looking at the output of `ceph status`, which should now include a mgr status line:

```
1. mgr active: $name
```

## Client authentication

The manager is a new daemon which requires new CephX capabilities. If you upgrade a cluster from an old version of Ceph, or use the default install/deploy tools, your admin client should get this capability automatically. If you use tooling from elsewhere, you may get EACCES errors when invoking certain ceph cluster commands. To fix that, add a “mgr allow \*” stanza to your client’s cephx capabilities by [Modifying User Capabilities](#).

## High availability

In general, you should set up a ceph-mgr on each of the hosts running a ceph-mon daemon to achieve the same level of availability.

By default, whichever ceph-mgr instance comes up first will be made active by the monitors, and the others will be standbys. There is no requirement for quorum among the ceph-mgr daemons.

If the active daemon fails to send a beacon to the monitors for more than `mon mgr beacon grace` (default 30s), then it will be replaced by a standby.

If you want to pre-empt failover, you can explicitly mark a ceph-mgr daemon as failed using `ceph mgr fail <mgr name>`.

## Using modules

Use the command `ceph mgr module ls` to see which modules are available, and which are currently enabled. Enable or disable modules using the commands `ceph mgr module enable <module>` and `ceph mgr module disable <module>` respectively.

If a module is *enabled* then the active ceph-mgr daemon will load and execute it. In the case of modules that provide a service, such as an HTTP server, the module may publish its address when it is loaded. To see the addresses of such modules, use the command `ceph mgr services`.

Some modules may also implement a special standby mode which runs on standby ceph-mgr daemons as well as the active daemon. This enables modules that provide services to redirect their clients to the active daemon, if the client tries to connect to a standby.

Consult the documentation pages for individual manager modules for more information about what functionality each module provides.

Here is an example of enabling the [Dashboard](#) module:

```
1. $ ceph mgr module ls
2. {
3.     "enabled_modules": [
4.         "restful",
5.         "status"
6.     ],
7.     "disabled_modules": [
8.         "dashboard"
9.     ]
10. }
11.
12. $ ceph mgr module enable dashboard
13. $ ceph mgr module ls
14. {
15.     "enabled_modules": [
16.         "restful",
17.         "status",
18.         "dashboard"
19.     ],
20.     "disabled_modules": [
21.     ]
22. }
23.
24. $ ceph mgr services
25. {
26.     "dashboard": "http://myserver.com:7789/",
```

```
27.      "restful": "https://myserver.com:8789/"
28. }
```

The first time the cluster starts, it uses the `mgr_initial_modules` setting to override which modules to enable. However, this setting is ignored through the rest of the lifetime of the cluster: only use it for bootstrapping. For example, before starting your monitor daemons for the first time, you might add a section like this to your `ceph.conf`:

```
1. [mon]
2.   mgr initial modules = dashboard balancer
```

## Calling module commands

Where a module implements command line hooks, the commands will be accessible as ordinary Ceph commands. Ceph will automatically incorporate module commands into the standard CLI interface and route them appropriately to the module.:

```
1. ceph <command | help>
```

## Configuration

`mgr module path`

Description

Path to load modules from

Type

String

Default

`"<library dir>/mgr"`

`mgr data`

Description

Path to load daemon data (such as keyring)

Type

String

Default

`"/var/lib/ceph/mgr/$cluster-$id"`

```
mgr tick period
```

#### Description

How many seconds between mgr beacons to monitors, and other periodic checks.

#### Type

Integer

#### Default

5

```
mon mgr beacon grace
```

#### Description

How long after last beacon should a mgr be considered failed

#### Type

Integer

#### Default

30

# ceph-mgr module developer's guide

## Warning

This is developer documentation, describing Ceph internals that are only relevant to people writing ceph-mgr modules.

## Creating a module

In pybind/mgr/, create a python module. Within your module, create a class that inherits from `MgrModule`. For ceph-mgr to detect your module, your directory must contain a file called `module.py`.

The most important methods to override are:

- a `serve` member function for server-type modules. This function should block forever.
- a `notify` member function if your module needs to take action when new cluster data is available.
- a `handle_command` member function if your module exposes CLI commands.

Some modules interface with external orchestrators to deploy Ceph services. These also inherit from `Orchestrator`, which adds additional methods to the base `MgrModule` class. See [Orchestrator modules](#) for more on creating these modules.

## Installing a module

Once your module is present in the location set by the `mgr module path` configuration setting, you can enable it via the `ceph mgr module enable` command:

```
1. ceph mgr module enable mymodule
```

Note that the `MgrModule` interface is not stable, so any modules maintained outside of the Ceph tree are liable to break when run against any newer or older versions of Ceph.

## Logging

Logging in Ceph manager modules is done as in any other Python program. Just import the `logging` package and get a logger instance with the `logging.getLogger` function.

Each module has a `log_level` option that specifies the current Python logging level of the module. To change or query the logging level of the module use the following Ceph

commands:

```
1. ceph config get mgr mgr/<module_name>/log_level
2. ceph config set mgr mgr/<module_name>/log_level <info|debug|critical|error|warning|>
```

The logging level used upon the module's start is determined by the current logging level of the mgr daemon, unless if the `log_level` option was previously set with the `config set ...` command. The mgr daemon logging level is mapped to the module python logging level as follows:

- `<= 0` is CRITICAL
- `<= 1` is WARNING
- `<= 4` is INFO
- `<= +inf` is DEBUG

We can unset the module log level and fallback to the mgr daemon logging level by running the following command:

```
1. ceph config set mgr mgr/<module_name>/log_level ''
```

By default, modules' logging messages are processed by the Ceph logging layer where they will be recorded in the mgr daemon's log file. But it's also possible to send a module's logging message to its own file.

The module's log file will be located in the same directory as the mgr daemon's log file with the following name pattern:

```
1. <mgr_daemon_log_file_name>.<module_name>.log
```

To enable the file logging on a module use the following command:

```
1. ceph config set mgr mgr/<module_name>/log_to_file true
```

When the module's file logging is enabled, module's logging messages stop being written to the mgr daemon's log file and are only written to the module's log file.

It's also possible to check the status and disable the file logging with the following commands:

```
1. ceph config get mgr mgr/<module_name>/log_to_file
2. ceph config set mgr mgr/<module_name>/log_to_file false
```

## Exposing commands

Set the `COMMANDS` class attribute of your module to a list of dicts like this:

```

1. COMMANDS = [
2.     {
3.         "cmd": "foobar name=myarg,type=CephString",
4.         "desc": "Do something awesome",
5.         "perm": "rw",
6.         # optional:
7.         "poll": "true"
8.     }
9. ]

```

The `cmd` part of each entry is parsed in the same way as internal Ceph mon and admin socket commands (see `mon/MonCommands.h` in the Ceph source for examples). Note that the “poll” field is optional, and is set to `False` by default; this indicates to the `ceph` CLI that it should call this command repeatedly and output results (see `ceph -h` and its `--period` option).

Each command is expected to return a tuple `(retval, stdout, stderr)`. `retval` is an integer representing a libc error code (e.g. `EINVAL`, `EPERM`, or `0` for no error), `stdout` is a string containing any non-error output, and `stderr` is a string containing any progress or error explanation output. Either or both of the two strings may be empty.

Implement the `handle_command` function to respond to the commands when they are sent:

```
MgrModule.^`handle_command (inbuf, cmd)
```

Called by `ceph-mgr` to request the plugin to handle one of the commands that it declared in `self.COMMANDS`

Return a status code, an output buffer, and an output string. The output buffer is for data results, the output string is for informative text.

- Parameters

- `inbuf (str)` – content of any “`-i <file>`” supplied to ceph cli
- `cmd (dict)` – from Ceph’s `cmdmap_t`

Returns

`HandleCommandResult` or a 3-tuple of `(int, str, str)`

## Configuration options

Modules can load and store configuration options using the `set_module_option` and `get_module_option` methods.

Note

Use `set_module_option` and `get_module_option` to manage user-visible configuration options that are not blobs (like certificates). If you want to persist module-internal data or binary configuration data consider using the [KV store](#).

You must declare your available configuration options in the `MODULE_OPTIONS` class attribute, like this:

```
1. MODULE_OPTIONS = [
2.     {
3.         "name": "my_option"
4.     }
5. ]
```

If you try to use `set_module_option` or `get_module_option` on options not declared in `MODULE_OPTIONS`, an exception will be raised.

You may choose to provide setter commands in your module to perform high level validation. Users can also modify configuration using the normal ceph config set command, where the configuration options for a mgr module are named like `mgr/<module name>/<option>`.

If a configuration option is different depending on which node the mgr is running on, then use *localized* configuration (`get_localized_module_option`, `set_localized_module_option`). This may be necessary for options such as what address to listen on. Localized options may also be set externally with `ceph config set`, where the key name is like `mgr/<module name>/<mgr id>/<option>`

If you need to load and store data (e.g. something larger, binary, or multiline), use the KV store instead of configuration options (see next section).

Hints for using config options:

- Reads are fast: ceph-mgr keeps a local in-memory copy, so in many cases you can just do a `get_module_option` every time you use a option, rather than copying it out into a variable.
- Writes block until the value is persisted (i.e. round trip to the monitor), but reads from another thread will see the new value immediately.
- If a user has used config set from the command line, then the new value will become visible to `get_module_option` immediately, although the mon->mgr update is asynchronous, so config set will return a fraction of a second before the new value is visible on the mgr.
- To delete a config value (i.e. revert to default), just pass `None` to `set_module_option`.

```
MgrModule.``get_module_option``(key, default=None)
```

Retrieve the value of a persistent configuration setting

- Parameters

- **key** (*str*) –
- **default** (*str*) –

Returns

*str*

```
MgrModule.``set_module_option (key, val)
```

Set the value of a persistent configuration setting

- Parameters

- key** (*str*) –

```
MgrModule.``get_localized_module_option (key, default=None)
```

Retrieve localized configuration for this ceph-mgr instance :param str key: :param str default: :return: str

```
MgrModule.``set_localized_module_option (key, val)
```

Set localized configuration for this ceph-mgr instance :param str key: :param str val: :return: str

## KV store

Modules have access to a private (per-module) key value store, which is implemented using the monitor’s “config-key” commands. Use the `set_store` and `get_store` methods to access the KV store from your module.

The KV store commands work in a similar way to the configuration commands. Reads are fast, operating from a local cache. Writes block on persistence and do a round trip to the monitor.

This data can be accessed from outside of ceph-mgr using the `ceph config-key [get|set]` commands. Key names follow the same conventions as configuration options. Note that any values updated from outside of ceph-mgr will not be seen by running modules until the next restart. Users should be discouraged from accessing module KV data externally – if it is necessary for users to populate data, modules should provide special commands to set the data via the module.

Use the `get_store_prefix` function to enumerate keys within a particular prefix (i.e. all keys starting with a particular substring).

```
MgrModule.``get_store (key, default=None)
```

Get a value from this module’s persistent key value store

```
MgrModule.` `set_store (key, val)
```

Set a value in this module's persistent key value store. If val is None, remove key from store

- Parameters

- **key** (str) -

- **val** (str) -

```
MgrModule.` `get_localized_store (key, default=None)
```

```
MgrModule.` `set_localized_store (key, val)
```

```
MgrModule.` `get_store_prefix (key_prefix)
```

Retrieve a dict of KV store keys to values, where the keys have the given prefix

- Parameters

- **key\_prefix** (str) -

- Returns

- str

## Accessing cluster data

---

Modules have access to the in-memory copies of the Ceph cluster's state that the mgr maintains. Accessor functions are exposed as members of MgrModule.

Calls that access the cluster or daemon state are generally going from Python into native C++ routines. There is some overhead to this, but much less than for example calling into a REST API or calling into an SQL database.

There are no consistency rules about access to cluster structures or daemon metadata. For example, an OSD might exist in OSDMap but have no metadata, or vice versa. On a healthy cluster these will be very rare transient states, but modules should be written to cope with the possibility.

Note that these accessors must not be called in the modules `__init__` function. This will result in a circular locking exception.

```
MgrModule.` `get (data_name)
```

Called by the plugin to fetch named cluster-wide objects from ceph-mgr.

- Parameters

- **data\_name** (str) - Valid things to fetch are osd\_crush\_map\_text, osd\_map,

```
osd_map_tree, osd_map_crush, config, mon_map, fs_map, osd_metadata, pg_summary,
io_rate, pg_dump, df, osd_stats, health, mon_status, devices, device <devid>,
pg_stats, pool_stats, pg_ready, osd_ping_times.
```

- Note:

All these structures have their own JSON representations: experiment or look at the C++ `dump()` methods to learn about them.

```
MgrModule.``get_server``(hostname)
```

Called by the plugin to fetch metadata about a particular hostname from ceph-mgr.

This is information that ceph-mgr has gleaned from the daemon metadata reported by daemons running on a particular server.

- Parameters

**hostname** – a hostname

```
MgrModule.``list_servers``()
```

Like `get_server`, but gives information about all servers (i.e. all unique hostnames that have been mentioned in daemon metadata)

- Returns

a list of information about all servers

Return type

list

```
MgrModule.``get_metadata``(svc_type, svc_id, default=None)
```

Fetch the daemon metadata for a particular service.

ceph-mgr fetches metadata asynchronously, so are windows of time during addition/removal of services where the metadata is not available to modules. `None` is returned if no metadata is available.

- Parameters

- **svc\_type** (*str*) – service type (e.g., ‘mds’, ‘osd’, ‘mon’)
- **svc\_id** (*str*) – service id. convert OSD integer IDs to strings when calling this

Return type

`dict`, or `None` if no metadata found

```
MgrModule.``get_daemon_status``(svc_type, svc_id)
```

Fetch the latest status for a particular service daemon.

This method may return `None` if no status information is available, for example because the daemon hasn't fully started yet.

- Parameters

- **svc\_type** – string (e.g., ‘rgw’)
- **svc\_id** – string

Returns

dict, or `None` if the service is not found

```
MgrModule.^`get_perf_schema` (svc_type, svc_name)
```

Called by the plugin to fetch perf counter schema info. `svc_name` can be `nullptr`, as can `svc_type`, in which case they are wildcards

- Parameters

- **svc\_type** (`str`) –
- **svc\_name** (`str`) –

Returns

list of dicts describing the counters requested

```
MgrModule.^`get_counter` (svc_type, svc_name, path)
```

Called by the plugin to fetch the latest performance counter data for a particular counter on a particular service.

- Parameters

- **svc\_type** (`str`) –
- **svc\_name** (`str`) –
- **path** (`str`) – a period-separated concatenation of the subsystem and the counter name, for example “mds.inodes”.

Returns

A list of two-tuples of (`timestamp`, `value`) is returned. This may be empty if no data is available.

```
MgrModule.^`get_mgr_id` ()
```

Retrieve the name of the manager daemon where this plugin is currently being executed (i.e. the active manager).

- Returns

str

## Exposing health checks

Modules can raise first class Ceph health checks, which will be reported in the output of `ceph status` and in other places that report on the cluster's health.

If you use `set_health_checks` to report a problem, be sure to call it again with an empty dict to clear your health check when the problem goes away.

```
MgrModule.``set_health_checks`` (checks)
```

Set the module's current map of health checks. Argument is a dict of check names to info, in this form:

```
1. {
2.     'CHECK_FOO': {
3.         'severity': 'warning',           # or 'error'
4.         'summary': 'summary string',
5.         'count': 4,                   # quantify badness
6.         'detail': [ 'list', 'of', 'detail', 'strings' ],
7.     },
8.     'CHECK_BAR': {
9.         'severity': 'error',
10.        'summary': 'bars are bad',
11.        'detail': [ 'too hard' ],
12.    },
13. }
```

- Parameters

`list` – dict of health check dicts

## What if the mons are down?

The manager daemon gets much of its state (such as the cluster maps) from the monitor. If the monitor cluster is inaccessible, whichever manager was active will continue to run, with the latest state it saw still in memory.

However, if you are creating a module that shows the cluster state to the user then you may well not want to mislead them by showing them that out of date state.

To check if the manager daemon currently has a connection to the monitor cluster, use this function:

```
MgrModule.``have_mon_connection`` ()
```

Check whether this ceph-mgr daemon has an open connection to a monitor. If it doesn't, then it's likely that the information we have about the cluster is out of date, and/or the monitor cluster is down.

## Reporting if your module cannot run

If your module cannot be run for any reason (such as a missing dependency), then you can report that by implementing the `can_run` function.

```
static MgrModule.``can_run``()
```

Implement this function to report whether the module's dependencies are met. For example, if the module needs to import a particular dependency to work, then use a try/except around the import at file scope, and then report here if the import failed.

This will be called in a blocking way from the C++ code, so do not do any I/O that could block in this function.

```
:return a 2-tuple consisting of a boolean and explanatory string
```

Note that this will only work properly if your module can always be imported: if you are importing a dependency that may be absent, then do it in a try/except block so that your module can be loaded far enough to use `can_run` even if the dependency is absent.

## Sending commands

A non-blocking facility is provided for sending monitor commands to the cluster.

```
MgrModule.``send_command``(\args, **kwargs*)
```

Called by the plugin to send a command to the mon cluster.

- Parameters

- **result** (`CommandResult`) – an instance of the `CommandResult` class, defined in the same module as `MgrModule`. This acts as a completion and stores the output of the command. Use `CommandResult.wait()` if you want to block on completion.
- **svc\_type** (`str`) –
- **svc\_id** (`str`) –
- **command** (`str`) – a JSON-serialized command. This uses the same format as the ceph command line, which is a dictionary of command arguments, with the extra `prefix` key containing the command name itself. Consult `MonCommands.h` for available commands and their expected arguments.
- **tag** (`str`) – used for nonblocking operation: when a command completes, the

`notify()` callback on the `MgrModule` instance is triggered, with `notify_type` set to “command”, and `notify_id` set to the tag of the command.

## Receiving notifications

---

The manager daemon calls the `notify` function on all active modules when certain important pieces of cluster state are updated, such as the cluster maps.

The actual data is not passed into this function, rather it is a cue for the module to go and read the relevant structure if it is interested. Most modules ignore most types of notification: to ignore a notification simply return from this function without doing anything.

```
MgrModule.``notify (notify_type, notify_id)
```

Called by the ceph-mgr service to notify the Python plugin that new state is available.

- Parameters

- `notify_type` – string indicating what kind of notification, such as `osd_map`, `mon_map`, `fs_map`, `mon_status`, `health`, `pg_summary`, `command`, `service_map`
- `notify_id` – string (may be empty) that optionally specifies which entity is being notified about. With “command” notifications this is set to the tag `from send_command`.

## Accessing RADOS or CephFS

---

If you want to use the librados python API to access data stored in the Ceph cluster, you can access the `rados` attribute of your `MgrModule` instance. This is an instance of `rados.Rados` which has been constructed for you using the existing Ceph context (an internal detail of the C++ Ceph code) of the mgr daemon.

Always use this specially constructed librados instance instead of constructing one by hand.

Similarly, if you are using libcephfs to access the file system, then use the libcephfs `create_with_rados` to construct it from the `MgrModule.rados` librados instance, and thereby inherit the correct context.

Remember that your module may be running while other parts of the cluster are down: do not assume that librados or libcephfs calls will return promptly – consider whether to use timeouts or to block if the rest of the cluster is not fully available.

## Implementing standby mode

---

For some modules, it is useful to run on standby manager daemons as well as on the active daemon. For example, an HTTP server can usefully serve HTTP redirect responses from the standby managers so that the user can point his browser at any of the manager daemons without having to worry about which one is active.

Standby manager daemons look for a subclass of `StandbyModule` in each module. If the class is not found then the module is not used at all on standby daemons. If the class is found, then its `serve` method is called. Implementations of `StandbyModule` must inherit from `mgr_module.MgrStandbyModule`.

The interface of `MgrStandbyModule` is much restricted compared to `MgrModule` – none of the Ceph cluster state is available to the module. `serve` and `shutdown` methods are used in the same way as a normal module class. The `get_active_uri` method enables the standby module to discover the address of its active peer in order to make redirects. See the `MgrStandbyModule` definition in the Ceph source code for the full list of methods.

For an example of how to use this interface, look at the source code of the `dashboard` module.

## Communicating between modules

Modules can invoke member functions of other modules.

```
MgrModule.``remote (module_name, method_name, \args, **kwargs*)
```

Invoke a method on another module. All arguments, and the return value from the other module must be serializable.

**Limitation:** Do not import any modules within the called method. Otherwise you will get an error in Python 2:

```
1. RuntimeError('cannot unmarshal code objects in restricted execution mode',)
```

- Parameters

- **module\_name** – Name of other module. If module isn't loaded, an `ImportError` exception is raised.
- **method\_name** – Method name. If it does not exist, a `NameError` exception is raised.
- **args** – Argument tuple
- **kwargs** – Keyword argument dict

**Raises**

- **RuntimeError** – Any error raised within the method is converted to a

## RuntimeError

- **ImportError** – No such module

Be sure to handle `ImportError` to deal with the case that the desired module is not enabled.

If the remote method raises a python exception, this will be converted to a `RuntimeError` on the calling side, where the message string describes the exception that was originally thrown. If your logic intends to handle certain errors cleanly, it is better to modify the remote method to return an error value instead of raising an exception.

At time of writing, inter-module calls are implemented without copies or serialization, so when you return a python object, you're returning a reference to that object to the calling module. It is recommended *not* to rely on this reference passing, as in future the implementation may change to serialize arguments and return values.

## Shutting down cleanly

---

If a module implements the `serve()` method, it should also implement the `shutdown()` method to shutdown cleanly: misbehaving modules may otherwise prevent clean shutdown of ceph-mgr.

## Limitations

---

It is not possible to call back into C++ code from a module's `__init__()` method. For example calling `self.get_module_option()` at this point will result in an assertion failure in ceph-mgr. For modules that implement the `serve()` method, it usually makes sense to do most initialization inside that method instead.

## Is something missing?

---

The ceph-mgr python interface is not set in stone. If you have a need that is not satisfied by the current interface, please bring it up on the ceph-devel mailing list. While it is desired to avoid bloating the interface, it is not generally very hard to expose existing data to the Python code when there is a good reason.

# ceph-mgr orchestrator modules

## Warning

This is developer documentation, describing Ceph internals that are only relevant to people writing ceph-mgr orchestrator modules.

In this context, *orchestrator* refers to some external service that provides the ability to discover devices and create Ceph services. This includes external projects such as Rook.

An *orchestrator module* is a ceph-mgr module ([ceph-mgr module developer's guide](#)) which implements common management operations using a particular orchestrator.

Orchestrator modules subclass the `Orchestrator` class: this class is an interface, it only provides method definitions to be implemented by subclasses. The purpose of defining this common interface for different orchestrators is to enable common UI code, such as the dashboard, to work with various different backends.

```
digraph G { subgraph cluster_1 { volumes [label="mgr/volumes"] rook [label="mgr/rook"] dashboard [label="mgr/dashboard"] orchestrator_cli [label="mgr/orchestrator"] orchestrator [label="Orchestrator Interface"] cephadm [label="mgr/cephadm"] label = "ceph-mgr"; } volumes -> orchestrator dashboard -> orchestrator orchestrator_cli -> orchestrator orchestrator -> rook -> rook_io orchestrator -> cephadm rook_io [label="Rook"] rankdir="TB"; }
```

Behind all the abstraction, the purpose of orchestrator modules is simple: enable Ceph to do things like discover available hardware, create and destroy OSDs, and run MDS and RGW services.

A tutorial is not included here: for full and concrete examples, see the existing implemented orchestrator modules in the Ceph source tree.

## Glossary

### Stateful service

a daemon that uses local storage, such as OSD or mon.

### Stateless service

a daemon that doesn't use any local storage, such as an MDS, RGW, nfs-ganesha, iSCSI gateway.

### Label

arbitrary string tags that may be applied by administrators to hosts. Typically

administrators use labels to indicate which hosts should run which kinds of service. Labels are advisory (from human input) and do not guarantee that hosts have particular physical capabilities.

Drive group

collection of block devices with common/shared OSD formatting (typically one or more SSDs acting as journals/dbs for a group of HDDs).

Placement

choice of which host is used to run a service.

## Key Concepts

---

The underlying orchestrator remains the source of truth for information about whether a service is running, what is running where, which hosts are available, etc. Orchestrator modules should avoid taking any internal copies of this information, and read it directly from the orchestrator backend as much as possible.

Bootstrapping hosts and adding them to the underlying orchestration system is outside the scope of Ceph's orchestrator interface. Ceph can only work on hosts when the orchestrator is already aware of them.

Calls to orchestrator modules are all asynchronous, and return *completion* objects (see below) rather than returning values immediately.

Where possible, placement of stateless services should be left up to the orchestrator.

## Completions and batching

---

All methods that read or modify the state of the system can potentially be long running. To handle that, all such methods return a *Completion* object. Orchestrator modules must implement the *process* method: this takes a list of completions, and is responsible for checking if they're finished, and advancing the underlying operations as needed.

Each orchestrator module implements its own underlying mechanisms for completions. This might involve running the underlying operations in threads, or batching the operations up before later executing in one go in the background. If implementing such a batching pattern, the module would do no work on any operation until it appeared in a list of completions passed into *process*.

Some operations need to show a progress. Those operations need to add a *ProgressReference* to the completion. At some point, the progress reference becomes *effective*, meaning that the operation has really happened (e.g. a service has actually been started).

```
Orchestrator.``process (completions)
```

Given a list of Completion instances, process any which are incomplete.

Callers should inspect the detail of each completion to identify partial completion/progress information, and present that information to the user.

This method should not block, as this would make it slow to query a status, while other long running operations are in progress.

- Return type

`None`

```
class orchestrator.``Completion (_first_promise=None, value=<object object>,
on_complete=None, name=None)
```

Combines multiple promises into one overall operation.

Completions are composable by being able to call one completion from another completion. I.e. making them re-usable using Promises E.g.:

```
1. >>>
2. ... return Orchestrator().get_hosts().then(self._create_osd)
```

where `get_hosts` returns a Completion of list of hosts and `_create_osd` takes a list of hosts.

The concept behind this is to store the computation steps explicit and then explicitly evaluate the chain:

```
1. >>>
2. ... p = Completion(on_complete=lambda x: x*x).then(on_complete=lambda x: str(x))
3. ... p.finalize(2)
4. ... assert p.result = "4"
```

or graphically:

```
1. +-----+      +-----+
2. |           | then |           |
3. | lambda x: x*x | ---> | lambda x: str(x)|
4. |           |           |           |
5. +-----+      +-----+
```

- `fail` (*e*)

Sets the whole completion to be failed with this exception and end the evaluation.

- *property* `has_result`

Has the operation already a result?

For Write operations, it can already have a result, if the orchestrator's configuration is persistently written. Typically this would indicate that an update had been written to a manifest, but that the update had not necessarily been pushed out to the cluster.

- Return type

`bool`

Returns

- *property* `is_errored`

Has the completion failed. Default implementation looks for `self.exception`. Can be overwritten.

- Return type

`bool`

- *property* `is_finished`

Could the external operation be deemed as complete, or should we wait? We must wait for a read operation only if it is not complete.

- Return type

`bool`

- *property* `needs_result`

Could the external operation be deemed as complete, or should we wait? We must wait for a read operation only if it is not complete.

- Return type

`bool`

- *property* `progress_reference`

`ProgressReference`. Marks this completion as a write completeion.

- Return type

`Optional [ ProgressReference ]`

- *property* `result`

The result of the operation that we were waited for. Only valid after calling `Orchestrator.process()` on this completion.

- Return type

~T

- `result_str ()`

Force a string.

- Return type

`str`

```
class orchestrator.``ProgressReference`` (message, mgr, completion=None)
```

- `completion : Optional[Callable[], Completion]`

The completion can already have a result, before the write operation is effective. `progress == 1` means, the services are created / removed.

- `property progress`

if a orchestrator module can provide a more detailed progress information, it needs to also call `progress.update()`.

## Error Handling

---

The main goal of error handling within orchestrator modules is to provide debug information to assist users when dealing with deployment errors.

```
class orchestrator.``OrchestratorError`` (msg, errno=- 22, event_kind_subject=None)
```

General orchestrator specific error.

Used for deployment, configuration or user errors.

It's not intended for programming errors or orchestrator internal errors.

```
class orchestrator.``NoOrchestrator`` (msg='No orchestrator configured (try `ceph orch set backend`)')
```

No orchestrator in configured.

```
class orchestrator.``OrchestratorValidationError`` (msg, errno=- 22, event_kind_subject=None)
```

Raised when an orchestrator doesn't support a specific feature.

In detail, orchestrators need to explicitly deal with different kinds of errors:

1. No orchestrator configured

See `NoOrchestrator`.

2. An orchestrator doesn't implement a specific method.

For example, an Orchestrator doesn't support `add_host`.

In this case, a `NotImplementedError` is raised.

### 3. Missing features within implemented methods.

E.g. optional parameters to a command that are not supported by the backend (e.g. the hosts field in `Orchestrator.apply_mons()` command with the rook backend).

See `OrchestratorValidationError`.

### 4. Input validation errors

The `orchestrator` module and other calling modules are supposed to provide meaningful error messages.

See `OrchestratorValidationError`.

### 5. Errors when actually executing commands

The resulting Completion should contain an error string that assists in understanding the problem. In addition, `Completion.is_errorred()` is set to `True`

### 6. Invalid configuration in the orchestrator modules

This can be tackled similar to 5.

All other errors are unexpected orchestrator issues and thus should raise an exception that are then logged into the mgr log file. If there is a completion object at that point, `Completion.result()` may contain an error message.

## Excluded functionality

- Ceph's orchestrator interface is not a general purpose framework for managing linux servers - it is deliberately constrained to manage the Ceph cluster's services only.
- Multipathed storage is not handled (multipathing is unnecessary for Ceph clusters). Each drive is assumed to be visible only on a single host.

## Host management

`Orchestrator.``add_host``(host_spec)`

Add a host to the orchestrator inventory.

- Parameters

`host` – hostname

## Return type

```
Completion [ str ]
```

```
Orchestrator.``remove_host`` (host)
```

Remove a host from the orchestrator inventory.

- Parameters

- host** ( str ) – hostname

## Return type

```
Completion [ str ]
```

```
Orchestrator.``get_hosts`` ()
```

Report the hosts in the cluster.

- Return type

```
Completion [ List [ HostSpec ] ]
```

## Returns

list of HostSpec

```
Orchestrator.``update_host_addr`` (host, addr)
```

Update a host's address

- Parameters

- **host** ( str ) – hostname
- **addr** ( str ) – address (dns name or IP)

## Return type

```
Completion [ str ]
```

```
Orchestrator.``add_host_label`` (host, label)
```

Add a host label

- Return type

```
Completion [ str ]
```

```
Orchestrator.``remove_host_label`` (host, label)
```

Remove a host label

- Return type

`Completion [ str ]`

```
class orchestrator.``HostSpec`` (hostname, addr=None, labels=None, status=None)
```

Information about hosts. Like e.g. `kubectl get nodes`

## Devices

---

`Orchestrator.``get_inventory`` (host_filter=None, refresh=False)`

Returns something that was created by ceph-volume inventory.

- Return type

`Completion [ List [ InventoryHost ] ]`

Returns

list of `InventoryHost`

```
class orchestrator.``InventoryFilter`` (labels=None, hosts=None)
```

When fetching inventory, use this filter to avoid unnecessarily scanning the whole estate.

- Typical use: filter by host when presenting UI workflow for configuring a particular server. filter by label when not all of estate is Ceph servers, and we want to only learn about the Ceph servers. filter by label when we are interested particularly in e.g. OSD servers.

```
class ceph.deployment.inventory.``Devices`` (devices)
```

A container for Device instances with reporting

```
class ceph.deployment.inventory.``Device`` (path, sys_api=None, available=None, rejected_reasons=None, lvs=None, device_id=None, lsm_data=None)
```

## Placement

---

A [Placement Specification](#) defines the placement of daemons of a specific service.

In general, stateless services do not require any specific placement rules as they can run anywhere that sufficient system resources are available. However, some orchestrators may not include the functionality to choose a location in this way. Optionally, you can specify a location when creating a stateless service.

```
class ceph.deployment.service_spec.``PlacementSpec`` (label=None, hosts=None, count=None,
```

```
host_pattern=None)
```

For APIs that need to specify a host subset

- `classmethod from_string(arg)`

A single integer is parsed as a count: >>> PlacementSpec.from\_string('3')  
PlacementSpec(count=3)

A list of names is parsed as host specifications: >>>  
PlacementSpec.from\_string('host1 host2') PlacementSpec(hosts=[HostPlacementSpec(hostname='host1', network='', name=''), HostPlacementSpec(hostname='host2', network='', name='')])

You can also prefix the hosts with a count as follows: >>>  
PlacementSpec.from\_string('2 host1 host2') PlacementSpec(count=2, hosts=[HostPlacementSpec(hostname='host1', network='', name=''), HostPlacementSpec(hostname='host2', network='', name='')])

You can specify labels using label:<label> >>>  
PlacementSpec.from\_string('label:mon') PlacementSpec(label='mon')

Labels also support a count: >>> PlacementSpec.from\_string('3 label:mon')  
PlacementSpec(count=3, label='mon')

fnmatch is also supported: >>> PlacementSpec.from\_string('data[1-3]')  
PlacementSpec(host\_pattern='data[1-3]')

```
1. >>> PlacementSpec.from_string(None)
2. PlacementSpec()
```

- Return type

`PlacementSpec`

- `host_pattern : Optional[str]`

fnmatch patterns to select hosts. Can also be a single host.

- `pretty_str()`

```
1. >>>
2. ... ps = PlacementSpec(...) # For all placement specs:
3. ... PlacementSpec.from_string(ps.pretty_str()) == ps
```

## Services

```
class orchestrator.`ServiceDescription` (spec, container_image_id=None,
container_image_name=None, rados_config_location=None, service_url=None,
```

```
last_refresh=None, created=None, size=0, running=0, events=None)
```

For responding to queries about the status of a particular service, stateful or stateless.

This is not about health or performance monitoring of services: it's about letting the orchestrator tell Ceph whether and where a service is scheduled in the cluster. When an orchestrator tells Ceph "it's running on host123", that's not a promise that the process is literally up this second, it's a description of where the orchestrator has decided the service should run.

```
class ceph.deployment.service_spec.``ServiceSpec (service_type, service_id=None, placement=None,
count=None, unmanaged=False, preview_only=False)
```

Details of service creation.

Request to the orchestrator for a cluster of daemons such as MDS, RGW, iscsi gateway, MONs, MGRs, Prometheus

This structure is supposed to be enough information to start the services.

```
Orchestrator.``describe_service (service_type=None, service_name=None, refresh=False)
```

Describe a service (of any kind) that is already configured in the orchestrator. For example, when viewing an OSD in the dashboard we might like to also display information about the orchestrator's view of the service (like the kubernetes pod ID).

When viewing a CephFS filesystem in the dashboard, we would use this to display the pods being currently run for MDS daemons.

- Return type

```
Completion [ List [ ServiceDescription ] ]
```

Returns

list of ServiceDescription objects.

```
Orchestrator.``service_action (action, service_name)
```

Perform an action (start/stop/reload) on a service (i.e., all daemons providing the logical service).

- Parameters

- **action** ( `str` ) – one of "start", "stop", "restart", "redeploy", "reconfig"
- **service\_name** ( `str` ) – service\_type + '.' + service\_id (e.g. "mon", "mgr", "mds.mycephfs", "rgw.realm.zone", ...)

Return type

## Completion

```
Orchestrator.``remove_service`` (service_name)
```

Remove a service (a collection of daemons).

- Return type

```
Completion [ str ]
```

Returns

None

## Daemons

---

```
Orchestrator.``list_daemons`` (service_name=None, daemon_type=None, daemon_id=None, host=None, refresh=False)
```

Describe a daemon (of any kind) that is already configured in the orchestrator.

- Return type

```
Completion [ List [ DaemonDescription ] ]
```

Returns

list of DaemonDescription objects.

```
Orchestrator.``remove_daemons`` (names)
```

Remove specific daemon(s).

- Return type

```
Completion [ List [ str ] ]
```

Returns

None

```
Orchestrator.``daemon_action`` (action, daemon_name, image=None)
```

Perform an action (start/stop/reload) on a daemon.

- Parameters

- **action** ( `str` ) – one of “start”, “stop”, “restart”, “redeploy”, “reconfig”
- **daemon\_name** ( `str` ) – name of daemon
- **image** ( `Optional` [ `str` ] ) – Container image when redeploying that daemon

- Return type

- [Completion](#)

## OSD management

---

```
Orchestrator.``create_osds``(drive_group)
```

Create one or more OSDs within a single Drive Group.

The principal argument here is the `drive_group` member of `OsdSpec`: other fields are advisory/extensible for any finer-grained OSD feature enablement (choice of backing store, compression/encryption, etc).

- Return type

- [Completion](#) [ [str](#) ]

```
Orchestrator.``blink_device_light``(ident_fault, on, locations)
```

Instructs the orchestrator to enable or disable either the ident or the fault LED.

- Parameters

- `ident_fault` ( [str](#) ) – either ["ident"](#) or ["fault"](#)
- `on` ( [bool](#) ) – [True](#) = on.
- `locations` ( [List](#) [ [DeviceLightLoc](#) ]) – See [orchestrator.DeviceLightLoc](#)

Return type

- [Completion](#) [ [List](#) [ [str](#) ]]

```
class orchestrator.``DeviceLightLoc``(host, dev, path)
```

Describes a specific device on a specific host. Used for enabling or disabling LEDs on devices.

hostname as in [orchestrator.Orchestrator.get\\_hosts\(\)](#)

- `device_id`: e.g. [ABC1234DEF567-1R1234\\_ABC8DE0Q](#).

See [ceph osd metadata | jq '.\[\].device\\_ids'](#)

## OSD Replacement

See [Replacing an OSD](#) for the underlying process.

Replacing OSDs is fundamentally a two-staged process, as users need to physically replace drives. The orchestrator therefore exposes this two-staged process.

Phase one is a call to `Orchestrator.remove_daemons()` with `destroy=True` in order to mark the OSD as destroyed.

Phase two is a call to `Orchestrator.create_osds()` with a Drive Group with `DriveGroupSpec.osd_id_claims` set to the destroyed OSD ids.

## Monitors

---

`Orchestrator.``add_mon (spec)`

Create mon daemon(s)

- Return type

`Completion [ List [ str ] ]`

`Orchestrator.``apply_mon (spec)`

Update mon cluster

- Return type

`Completion [ str ]`

## Stateless Services

---

`Orchestrator.``add_mgr (spec)`

Create mgr daemon(s)

- Return type

`Completion [ List [ str ] ]`

`Orchestrator.``apply_mgr (spec)`

Update mgr cluster

- Return type

`Completion [ str ]`

`Orchestrator.``add_mds (spec)`

Create MDS daemon(s)

- Return type

`Completion [ List [ str ] ]`

```
Orchestrator.``apply_mds``(spec)
```

Update MDS cluster

- Return type

```
Completion [ str ]
```

```
Orchestrator.``add_rbd_mirror``(spec)
```

Create rbd-mirror daemon(s)

- Return type

```
Completion [ List [ str ] ]
```

```
Orchestrator.``apply_rbd_mirror``(spec)
```

Update rbd-mirror cluster

- Return type

```
Completion [ str ]
```

```
class ceph.deployment.service_spec.``RGWSpec``(service_type='rgw', service_id=None, placement=None, rgw_realm=None, rgw_zone=None, subcluster=None, rgw_frontend_port=None, rgw_frontend_ssl_certificate=None, rgw_frontend_ssl_key=None, unmanaged=False, ssl=False, preview_only=False)
```

Settings to configure a (multisite) Ceph RGW

```
Orchestrator.``add_rgw``(spec)
```

Create RGW daemon(s)

- Return type

```
Completion [ List [ str ] ]
```

```
Orchestrator.``apply_rgw``(spec)
```

Update RGW cluster

- Return type

```
Completion [ str ]
```

```
class ceph.deployment.service_spec.``NFSServiceSpec``(service_type='nfs', service_id=None, pool=None, namespace=None, placement=None, unmanaged=False, preview_only=False)
```

```
Orchestrator.``add_nfs``(spec)
```

Create NFS daemon(s)

- Return type

```
Completion [ List [ str ] ]
```

```
Orchestrator.``apply_nfs``(spec)
```

Update NFS cluster

- Return type

```
Completion [ str ]
```

## Upgrades

---

```
Orchestrator.``upgrade_available``()
```

Report on what versions are available to upgrade to

- Return type

```
Completion
```

Returns

List of strings

```
Orchestrator.``upgrade_start``(image, version)
```

- Return type

```
Completion [ str ]
```

```
Orchestrator.``upgrade_status``()
```

If an upgrade is currently underway, report on where we are in the process, or if some error has occurred.

- Return type

```
Completion [ UpgradeStatusSpec ]
```

Returns

UpgradeStatusSpec instance

```
class orchestrator.``UpgradeStatusSpec``
```

## Utility

---

```
Orchestrator.``available``()
```

Report whether we can talk to the orchestrator. This is the place to give the user a

meaningful message if the orchestrator isn't running or can't be contacted.

This method may be called frequently (e.g. every page load to conditionally display a warning banner), so make sure it's not too expensive. It's okay to give a slightly stale status (e.g. based on a periodic background ping of the orchestrator) if that's necessary to make this method fast.

#### Note

True doesn't mean that the desired functionality is actually available in the orchestrator. I.e. this won't work as expected:

```
1. >>>
2. ... if OrchestratorClientMixin().available()[0]: # wrong.
3. ...     OrchestratorClientMixin().get_hosts()
```

- Return type

`Tuple [ bool , str ]`

Returns

two-tuple of boolean, string

`Orchestrator.``get_feature_set` ()`

Describes which methods this orchestrator implements

#### Note

True doesn't mean that the desired functionality is actually possible in the orchestrator. I.e. this won't work as expected:

```
1. >>>
2. ... api = OrchestratorClientMixin()
3. ... if api.get_feature_set()['get_hosts']['available']: # wrong.
4. ...     api.get_hosts()
```

It's better to ask for forgiveness instead:

```
1. >>>
2. ... try:
3. ...     OrchestratorClientMixin().get_hosts()
4. ... except (OrchestratorError, NotImplementedError):
5. ...     ...
```

- Returns

Dict of API method names to `{'available': True or False}`

# Client Modules

```
class orchestrator.``OrchestratorClientMixin``
```

A module that inherits from `OrchestratorClientMixin` can directly call all `Orchestrator` methods without manually calling `remote`.

Every interface method from `Orchestrator` is converted into a stub method that internally calls `OrchestratorClientMixin._oremote()`

```
1. >>> class MyModule(OrchestratorClientMixin):
2. ...     def func(self):
3. ...         completion = self.add_host('somehost') # calls `__oremote()`
4. ...         self._orchestrator_wait([completion])
5. ...         self.log.debug(completion.result)
```

## Note

`Orchestrator` implementations should not inherit from `OrchestratorClientMixin`. Reason is, that `OrchestratorClientMixin` magically redirects all methods to the “real” implementation of the orchestrator.

```
1. >>> import mgr_module
2. >>>
3. ... class MyImplementation(mgr_module.MgrModule, Orchestrator):
4. ...     def __init__(self, ...):
5. ...         self.orch_client = OrchestratorClientMixin()
6. ...         self.orch_client.set_mgr(self.mgr))
```

- `set_mgr (mgr)`

Useable in the Dashboard that uses a global `mgr`

- Return type

`None`

# Ceph RESTful API

## Introduction

The **Ceph RESTful API** (henceforth **Ceph API**) is provided by the [Ceph Dashboard](#) module. The Ceph API service is available at the same URL as the regular Ceph Dashboard, under the `/api` base path (please refer to [Host Name and Port](#)):

```
1. http://<server_addr>:<server_port>/api
```

or, if HTTPS is enabled (please refer to [SSL/TLS Support](#)):

```
1. https://<server_addr>:<ssl_server_port>/api
```

The Ceph API leverages the following standards:

- [HTTP 1.1](#) for API syntax and semantics,
- [JSON](#) for content encoding,
- [HTTP Content Negotiation](#) and [MIME](#) for versioning,
- [OAuth 2.0](#) and [JWT](#) for authentication and authorization.

### Warning

Some endpoints are still under active development, and should be carefully used since new Ceph releases could bring backward incompatible changes.

## Authentication and Authorization

Requests to the Ceph API pass through two access control checkpoints:

- **Authentication:** ensures that the request is performed on behalf of an existing and valid user account.
- **Authorization:** ensures that the previously authenticated user can in fact perform a specific action (create, read, update or delete) on the target endpoint.

So, prior to start consuming the Ceph API, a valid JSON Web Token (JWT) has to be obtained, and it may then be reused for subsequent requests. The `/api/auth` endpoint will provide the valid token:

```
1. $ curl -X POST "https://example.com:8443/api/auth" \
2.   -H "Accept: application/vnd.ceph.api.v1.0+json" \
3.   -H "Content-Type: application/json" \
```

```

4. -d '{"username": <username>, "password": <password>}'
5.
6. { "token": "<redacted_token>", ...}

```

The token obtained must be passed together with every API request in the [Authorization](#) HTTP header:

```
1. curl -H "Authorization: Bearer <token>" ...
```

Authentication and authorization can be further configured from the Ceph CLI, the Ceph-Dashboard UI and the Ceph API itself (please refer to [User and Role Management](#)).

## Versioning

One of the main goals of the Ceph API is to keep a stable interface. For this purpose, Ceph API is built upon the following principles:

- **Mandatory:** in order to avoid implicit defaults, all endpoints require an explicit default version (starting with [1.0](#) ).
- **Per-endpoint:** as this API wraps many different Ceph components, this allows for a finer-grained change control.
  - **Content/MIME Type:** the version expected from a specific endpoint is stated by the [Accept: application/vnd.ceph.api.v<major>.v<minor>+json](#) HTTP header. If the current Ceph API server is not able to address that specific major version, a [415 - Unsupported Media Type](#) response will be returned.
- **Semantic Versioning:** with a [major.minor](#) version:
  - Major changes are backward incompatible: they might result in non-additive changes to the request and/or response formats of a specific endpoint.
  - Minor changes are backward/forward compatible: they basically consists of additive changes to the request or response formats of a specific endpoint.

An example:

```

1. $ curl -X GET "https://example.com:8443/api/osd" \
2.   -H "Accept: application/vnd.ceph.api.v1.0+json" \
3.   -H "Authorization: Bearer <token>"

```

## Specification

### Auth

**POST** /api/auth

### Example request:

```

1. POST /api/auth HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "password": "string",
7.     "username": "string"
8. }
```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**POST** /api/auth/check

### Check token Authentication

- Query Parameters

- token (string) – Authentication Token (Required)

### Example request:

```

1. POST /api/auth/check?token=string HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "token": "string"
7. }
```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.

- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

`POST /api/auth/logout`

- Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## Cephfs

`GET /api/cephfs`

### Example request:

- ```
1. GET /api/cephfs HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET** /api/cephfs/{fs\_id}

- Parameters

- **fs\_id** (*string*) –

**Example request:**

```
1. GET /api/cephfs/{fs_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/cephfs/{fs\_id}/client/{client\_id}

- Parameters

- **fs\_id** (*string*) –
- **client\_id** (*string*) –

**Status Codes**

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/cephfs/{fs\_id}/clients

- Parameters

- **fs\_id** (string) –

### Example request:

```
1. GET /api/cephfs/{fs_id}/clients HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/cephfs/{fs_id}/get_root_directory`

The root directory that can't be fetched using ls\_dir (api). :param fs\_id: The filesystem identifier. :return: The root directory :rtype: dict

- Parameters

- **fs\_id** (string) –

### Example request:

```
1. GET /api/cephfs/{fs_id}/get_root_directory HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/cephfs/{fs_id}/ls_dir`

List directories of specified path. :param fs\_id: The filesystem identifier. :param path: The path where to start listing the directory content. Defaults to '' if not set. :type path: str | bytes :param depth: The number of steps to go down the directory tree. :type depth: int | str :return: The names of the directories below the specified path. :rtype: list

- Parameters

- **fs\_id** (*string*) –

Query Parameters

- **path** (*string*) –
- **depth** (*integer*) –

**Example request:**

```
1. GET /api/cephfs/{fs_id}/ls_dir HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

GET      /api/cephfs/{fs\_id}/mds\_counters

- Parameters

- **fs\_id** (*string*) –

Query Parameters

- **counters** (*integer*) –

**Example request:**

```
1. GET /api/cephfs/{fs_id}/mds_counters HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/cephfs/{fs\_id}/quota

## Get Cephfs Quotas of the specified path

Get the quotas of the specified path. :param fs\_id: The filesystem identifier. :param path: The path of the directory/file. :return: Returns a dictionary containing 'max\_bytes' and 'max\_files'. :rtype: dict

- Parameters

- **fs\_id (string)** – File System Identifier

Query Parameters

- **path (string)** – File System Path (Required)

**Example request:**

```
1. GET /api/cephfs/{fs_id}/quota?path=string HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/cephfs/{fs\_id}/quota

Set the quotas of the specified path. :param fs\_id: The filesystem identifier. :param path: The path of the directory/file. :param max\_bytes: The byte limit. :param max\_files: The file limit.

- Parameters

- **fs\_id (string)** –

**Example request:**

```

1. PUT /api/cephfs/{fs_id}/quota HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "max_bytes": "string",
7.   "max_files": "string",
8.   "path": "string"
9. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**`/api/cephfs/{fs_id}/snapshot`

Remove a snapshot. :param fs\_id: The filesystem identifier. :param path: The path of the directory. :param name: The name of the snapshot.

- Parameters

- **fs\_id (string)** –

## Query Parameters

- **path (string)** – (Required)
- **name (string)** – (Required)

## Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/cephfs/{fs\_id}/snapshot

Create a snapshot. :param fs\_id: The filesystem identifier. :param path: The path of the directory. :param name: The name of the snapshot. If not specified, a name using the current time in RFC3339 UTC format will be generated. :return: The name of the snapshot. :rtype: str

- Parameters

- **fs\_id** (*string*) –

**Example request:**

```

1. POST /api/cephfs/{fs_id}/snapshot HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "name": "string",
7.     "path": "string"
8. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/cephfs/{fs\_id}/tree

Remove a directory. :param fs\_id: The filesystem identifier. :param path: The path of the directory.

- Parameters

- **fs\_id** (*string*) –

## Query Parameters

- **path** (*string*) – (Required)

## Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**`/api/cephfs/{fs_id}/tree`

Create a directory. :param fs\_id: The filesystem identifier. :param path: The path of the directory.

### • Parameters

- **fs\_id** (*string*) –

### **Example request:**

```

1. POST /api/cephfs/{fs_id}/tree HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "path": "string"
7. }
```

### • Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

## ClusterConfiguration

GET /api/cluster\_conf

### Example request:

```
1. GET /api/cluster_conf HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

POST /api/cluster\_conf

### Example request:

```
1. POST /api/cluster_conf HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "name": "string",
7.     "value": "string"
8. }
```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.

- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/cluster\_conf

#### Example request:

```

1. PUT /api/cluster_conf HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "options": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/cluster\_conf/filter

#### Get Cluster Configuration by name

- Query Parameters

- **names (string)** – Config option names

#### Example request:

```

1. GET /api/cluster_conf/filter HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** `/api/cluster_conf/{name}`

- Parameters

- **name (string)** –

Query Parameters

- **section (string)** – (Required)

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** `/api/cluster_conf/{name}`

- Parameters

- **name (string)** –

**Example request:**

```
1. GET /api/cluster_conf/{name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## CrushRule

[GET](#) [/api/crush\\_rule](#)

### List Crush Rule Configuration

#### Example request:

- ```
1. GET /api/crush_rule HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

[POST](#) [/api/crush\\_rule](#)

#### Example request:

```
1. POST /api/crush_rule HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "device_class": "string",
7.   "failure_domain": "string",
8.   "name": "string",
9.   "root": "string"
10. }
```

- Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/crush\_rule/{name}

- Parameters

- **name (string)** –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/crush\_rule/{name}

- Parameters

- **name (string)** –

**Example request:**

```
1. GET /api/crush_rule/{name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## ErasureCodeProfile

GET [/api/erasure\\_code\\_profile](/api/erasure_code_profile)

### List Erasure Code Profile Information

#### Example request:

```
1. GET /api/erasure_code_profile HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

POST [/api/erasure\\_code\\_profile](/api/erasure_code_profile)

#### Example request:

```
1. POST /api/erasure_code_profile HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "name": "string"
7. }
```

- Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/erasure\_code\_profile/{name}

- Parameters

- **name (string)** –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/erasure\_code\_profile/{name}

- Parameters

- **name (string)** –

#### Example request:

```
1. GET /api/erasure_code_profile/{name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

## FeatureTogglesEndpoint

GET /api/feature\_toggles

### Get List Of Features

#### Example request:

```
1. GET /api/feature_toggles HTTP/1.1
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## Grafana

POST /api/grafana/dashboards

- Status Codes
  - 201 Created – Resource created.
  - 202 Accepted – Operation is still executing. Please check the task queue.
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

GET /api/grafana/url

### List Grafana URL Instance

**Example request:**

```
1. GET /api/grafana/url HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/grafana/validation/{params}`

- Parameters
  - **params (string)** –

**Example request:**

```
1. GET /api/grafana/validation/{params} HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Health

**GET**    `/api/health/full`

**Example request:**

```
1. GET /api/health/full HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

`GET /api/health/minimal`

### Get Cluster's minimal health report

#### Example request:

```
1. GET /api/health/minimal HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## Host

`GET /api/host`

### List Host Specifications

- Query Parameters

- sources (*string*) – Host Sources

#### Example request:

```

1. GET /api/host HTTP/1.1
2. Host: example.com

```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/host`

**Example request:**

```

1. POST /api/host HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "hostname": "string"
7. }

```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**DELETE**    `/api/host/{hostname}`

- Parameters

- **hostname** (*string*) –

## Status Codes

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [204 No Content](#) – Resource deleted.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET**    </api/host/{hostname}>

Get the specified host. :raises: cherrypy.HTTPError: If host not found.

- Parameters

- **hostname** (*string*) –

### Example request:

```
1. GET /api/host/{hostname} HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**PUT**    </api/host/{hostname}>

Update the specified host. Note, this is only supported when Ceph Orchestrator is enabled. :param hostname: The name of the host to be processed. :param labels: List of labels.

- Parameters

- **hostname** (*string*) –

**Example request:**

```

1. PUT /api/host/{hostname} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "labels": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/host/{hostname}/daemons`

- Parameters

- **hostname (string)** –

**Example request:**

```

1. GET /api/host/{hostname}/daemons HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/host/{hostname}/devices

- Parameters

- **hostname** (*string*) -

**Example request:**

```
1. GET /api/host/{hostname}/devices HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/host/{hostname}/identify\_device

Identify a device by switching on the device light for N seconds. :param hostname: The hostname of the device to process. :param device: The device identifier to process, e.g. `/dev/dm-0` or `ABC1234DEF567-1R1234_ABC8DE0Q`. :param duration: The duration in seconds how long the LED should flash.

- Parameters

- **hostname** (*string*) -

**Example request:**

```
1. POST /api/host/{hostname}/identify_device HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "device": "string",
7.     "duration": "string"
8. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/host/{hostname}/inventory

### Get inventory of a host

- Parameters

- **hostname** (*string*) – Hostname

Query Parameters

- **refresh** (*string*) – Trigger asynchronous refresh

**Example request:**

```
1. GET /api/host/{hostname}/inventory HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/host/{hostname}/smart

- Parameters

- **hostname** (*string*) –

**Example request:**

```
1. GET /api/host/{hostname}/smart HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

# ISCSI

**GET** /api/iscsi/discoveryauth

## Get Iscsi discoveryauth Details

### Example request:

```
1. GET /api/iscsi/discoveryauth HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/iscsi/discoveryauth

## Set Iscsi discoveryauth

- Query Parameters

- user (string) – Username (Required)
- password (string) – Password (Required)
- mutual\_user (string) – Mutual UserName (Required)
- mutual\_password (string) – Mutual Password (Required)

### Example request:

```
1. PUT /api/iscsi/discoveryauth?user=string&password=string&mutual_user=string&mutual_password=string HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "mutual_password": "string",
7.     "mutual_user": "string",
8.     "password": "string",
9.     "user": "string"
10. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## IscsiTarget

**GET**    `/api/iscsi/target`

### Example request:

- ```

1. GET /api/iscsi/target HTTP/1.1
2. Host: example.com

```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/iscsi/target`

### Example request:

- ```

1. POST /api/iscsi/target HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "acl_enabled": "string",
7.     "auth": "string",

```

```

8.   "clients": "string",
9.   "disks": "string",
10.  "groups": "string",
11.  "portals": "string",
12.  "target_controls": "string",
13.  "target_iqn": "string"
14. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/iscsi/target/{target\_iqn}

- Parameters

- **target\_iqn (string)** –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/iscsi/target/{target\_iqn}

- Parameters

- **target\_iqn (string)** –

#### Example request:

```

1. GET /api/iscsi/target/{target_iqn} HTTP/1.1
2. Host: example.com

```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/iscsi/target/{target\_iqn}

- Parameters

- target\_iqn (string) –

#### Example request:

```

1. PUT /api/iscsi/target/{target_iqn} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "acl_enabled": "string",
7.   "auth": "string",
8.   "clients": "string",
9.   "disks": "string",
10.  "groups": "string",
11.  "new_target_iqn": "string",
12.  "portals": "string",
13.  "target_controls": "string"
14. }

```

- Status Codes

- 200 OK – Resource updated.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Logs

GET /api/logs/all

### Display Logs Configuration

#### Example request:

```
1. GET /api/logs/all HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## MdsPerfCounter

GET /api/perf\_counters/mds/{service\_id}

- Parameters
  - **service\_id** (*string*) –

#### Example request:

```
1. GET /api/perf_counters/mds/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## MgrModule

**GET** /api/mgr/module

### List Mgr modules

Get the list of managed modules. :return: A list of objects with the fields ‘enabled’, ‘name’ and ‘options’.  
:rtype: list

#### Example request:

1. GET /api/mgr/module HTTP/1.1
2. Host: example.com

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/mgr/module/{module\_name}

Retrieve the values of the persistent configuration settings. :param module\_name: The name of the Ceph Mgr module.  
:type module\_name: str :return: The values of the module options. :rtype: dict

- Parameters

- **module\_name (string)** –

#### Example request:

1. GET /api/mgr/module/{module\_name} HTTP/1.1
2. Host: example.com

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**PUT** [/api/mgr/module/{module\\_name}](/api/mgr/module/{module_name})

Set the values of the persistent configuration settings. :param module\_name: The name of the Ceph Mgr module. :type module\_name: str :param config: The values of the module options to be stored. :type config: dict

- Parameters

- **module\_name** (*string*) –

**Example request:**

```

1. PUT /api/mgr/module/{module_name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "config": "string"
7. }
```

- Status Codes

- [200 OK](#) – Resource updated.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**POST** [/api/mgr/module/{module\\_name}/disable](/api/mgr/module/{module_name}/disable)

Disable the specified Ceph Mgr module. :param module\_name: The name of the Ceph Mgr module. :type module\_name: str

- Parameters

- **module\_name** (*string*) –

#### Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    [/api/mgr/module/{module\\_name}/enable](#)

Enable the specified Ceph Mgr module. :param module\_name: The name of the Ceph Mgr module. :type module\_name: str

- Parameters

- **module\_name** (*string*) –

#### Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    [/api/mgr/module/{module\\_name}/options](#)

Get the module options of the specified Ceph Mgr module. :param module\_name: The name of the Ceph Mgr module. :type module\_name: str :return: The module options as list of dicts. :rtype: list

- Parameters

- **module\_name** (*string*) –

#### Example request:

```

1. GET /api/mgr/module/{module_name}/options HTTP/1.1
2. Host: example.com

```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## MgrPerfCounter

`GET /api/perf_counters/mgr/{service_id}`

- Parameters

- `service_id` (string) –

### Example request:

```

1. GET /api/perf_counters/mgr/{service_id} HTTP/1.1
2. Host: example.com

```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## MonPerfCounter

`GET /api/perf_counters/mon/{service_id}`

- Parameters

- **service\_id** (*string*) –

#### Example request:

```
1. GET /api/perf_counters/mon/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Monitor

**GET**    </api/monitor>

#### Get Monitor Details

#### Example request:

```
1. GET /api/monitor HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## NFS-Ganesha

**GET**    </api/nfs-ganesha/daemon>

## List NFS-Ganesha daemons information

### Example request:

```
1. GET /api/nfs-ganesha/daemon HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/nfs-ganesha/export`

## List all NFS-Ganesha exports

### Example request:

```
1. GET /api/nfs-ganesha/export HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/nfs-ganesha/export`

## Creates a new NFS-Ganesha export

### Example request:

```
1. POST /api/nfs-ganesha/export HTTP/1.1
```

```

2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "access_type": "string",
7.     "clients": [
8.         {
9.             "access_type": "string",
10.            "addresses": [
11.                "string"
12.            ],
13.            "squash": "string"
14.        }
15.    ],
16.    "cluster_id": "string",
17.    "daemons": [
18.        "string"
19.    ],
20.    "fsal": {
21.        "filesystem": "string",
22.        "name": "string",
23.        "rgw_user_id": "string",
24.        "sec_label_xattr": "string",
25.        "user_id": "string"
26.    },
27.    "path": "string",
28.    "protocols": [
29.        1
30.    ],
31.    "pseudo": "string",
32.    "reload_daemons": true,
33.    "security_label": "string",
34.    "squash": "string",
35.    "tag": "string",
36.    "transports": [
37.        "string"
38.    ]
39. }

```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

**DELETE** /api/nfs-ganesha/export/{cluster\_id}/{export\_id}

### Deletes an NFS-Ganesha export

- Parameters

- **cluster\_id** (*string*) – Cluster identifier
- **export\_id** (*integer*) – Export ID

#### Query Parameters

- **reload\_daemons** (*boolean*) – Trigger reload of NFS-Ganesha daemons configuration

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/nfs-ganesha/export/{cluster\_id}/{export\_id}

### Get an NFS-Ganesha export

- Parameters

- **cluster\_id** (*string*) – Cluster identifier
- **export\_id** (*integer*) – Export ID

#### Example request:

```
1. GET /api/nfs-ganesha/export/{cluster_id}/{export_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

`PUT`

`/api/nfs-ganesha/export/{cluster_id}/{export_id}`

## Updates an NFS-Ganesha export

- Parameters

- **cluster\_id** (*string*) – Cluster identifier
- **export\_id** (*integer*) – Export ID

### Example request:

```

1. PUT /api/nfs-ganesha/export/{cluster_id}/{export_id} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "access_type": "string",
7.     "clients": [
8.         {
9.             "access_type": "string",
10.            "addresses": [
11.                "string"
12.            ],
13.            "squash": "string"
14.        }
15.    ],
16.    "daemons": [
17.        "string"
18.    ],
19.    "fsal": {
20.        "filesystem": "string",
21.        "name": "string",
22.        "rgw_user_id": "string",
23.        "sec_label_xattr": "string",
24.        "user_id": "string"
25.    },
26.    "path": "string",
27.    "protocols": [
28.        1
29.    ],
30.    "pseudo": "string",
31.    "reload_daemons": true,
32.    "security_label": "string",

```

```

33.     "squash": "string",
34.     "tag": "string",
35.     "transports": [
36.         "string"
37.     ]
38. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    </api/nfs-ganesha/status>

### Status of NFS-Ganesha management feature

#### Example request:

```

1. GET /api/nfs-ganesha/status HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## OSD

**GET**    </api/osd>

#### Example request:

```

1. GET /api/osd HTTP/1.1
2. Host: example.com

```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**POST**    **/api/osd**

**Example request:**

```

1. POST /api/osd HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "data": "string",
7.   "method": "string",
8.   "tracking_id": "string"
9. }

```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**GET**    **/api/osd/flags**

**Display OSD Flags**

**Example request:**

```
1. GET /api/osd/flags HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT**    **/api/osd/flags**

**Sets OSD flags for the entire cluster.**

The recovery\_deletes, sortbitwise and pglog\_hardlimit flags cannot be unset. purged\_snapshots cannot even be set. It is therefore required to at least include those four flags for a successful operation.

**Example request:**

```
1. PUT /api/osd/flags HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "flags": [
7.     "string"
8.   ]
9. }
```

- Status Codes
  - **200 OK** – Resource updated.
  - **202 Accepted** – Operation is still executing. Please check the task queue.
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

**GET** /api/osd/flags/individual

### Displays individual OSD flags

#### Example request:

```
1. GET /api/osd/flags/individual HTTP/1.1
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/osd/flags/individual

### Sets OSD flags for a subset of individual OSDs.

Updates flags (noout, noin, nodown, noup) for an individual subset of OSDs.

#### Example request:

```
1. PUT /api/osd/flags/individual HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "flags": {
7.     "nodown": true,
8.     "noin": true,
9.     "noout": true,
10.    "noup": true
11.   },
12.   "ids": [
13.     1
14.   ]
15. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

GET

/api/osd/safe\_to\_delete

- type ids

int|[int]

- Query Parameters

- **svc\_ids** (*string*) – (Required)

**Example request:**

1. GET /api/osd/safe\_to\_delete?svc\_ids=*string* HTTP/1.1
2. Host: example.com

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

GET

/api/osd/safe\_to\_destroy

**Check If OSD is Safe to Destroy**

- type ids

int|[int]

- Query Parameters

- **ids** (*string*) – OSD Service Identifier (Required)

**Example request:**

```
1. GET /api/osd/safe_to_destroy?ids=string HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**`/api/osd/{svc_id}`

- Parameters

- **svc\_id** (*string*) –

Query Parameters

- **preserve\_id** (*string*) –
- **force** (*string*) –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/osd/{svc\_id}

Returns collected data about an OSD.

- return

Returns the requested data.

- Parameters

- **svc\_id** (*string*) –

**Example request:**

```
1. GET /api/osd/{svc_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/osd/{svc\_id}

- Parameters

- **svc\_id** (*string*) –

**Example request:**

```
1. PUT /api/osd/{svc_id} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "device_class": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**POST**[/api/osd/{svc\\_id}/destroy](#)

Mark osd as being destroyed. Keeps the ID intact (allowing reuse), but removes cephx keys, config-key data and lockbox keys, rendering data permanently unreadable.

The osd must be marked down before being destroyed.

- Parameters

- **svc\_id** (*string*) –

#### Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET**[/api/osd/{svc\\_id}/devices](#)

- Parameters

- **svc\_id** (*string*) –

#### Example request:

1. GET /api/osd/{svc\_id}/devices HTTP/1.1
2. Host: example.com

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET**    [/api/osd/{svc\\_id}/histogram](/api/osd/{svc_id}/histogram)

- ```

• return

Returns the histogram data.

```

- Parameters

- **svc\_id** (*string*) –

**Example request:**

```

1. GET /api/osd/{svc_id}/histogram HTTP/1.1
2. Host: example.com

```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**PUT**    [/api/osd/{svc\\_id}/mark](/api/osd/{svc_id}/mark)

**Mark OSD flags (out, in, down, lost, ...)**

```

Note: osd must be marked down before marking lost.

```

- Parameters

- **svc\_id** (*string*) – SVC ID

**Example request:**

```

1. PUT /api/osd/{svc_id}/mark HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "action": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

POST

/api/osd/{svc\_id}/purge

Note: osd must be marked down before removal.

- Parameters

- **svc\_id (string)** –

## Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

POST

/api/osd/{svc\_id}/reweight

Reweights the OSD temporarily.

Note that 'ceph osd reweight' is not a persistent setting. When an OSD gets marked out, the osd weight will be set to 0. When it gets marked in again, the weight will be changed to 1.

Because of this 'ceph osd reweight' is a temporary solution. You should only use it to keep your cluster running while you're ordering more hardware.

- Craig Lewis (<http://lists.ceph.com/pipermail/ceph-users-ceph.com/2014-June/040967.html>)

- Parameters

- **svc\_id** (*string*) –

**Example request:**

```

1. POST /api/osd/{svc_id}/reweight HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "weight": "string"
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

POST

/api/osd/{svc\_id}/scrub

- Parameters

- **svc\_id** (*string*) –

Query Parameters

- **deep** (*boolean*) –

**Example request:**

```
1. POST /api/osd/{svc_id}/scrub HTTP/1.1
```

```

2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "deep": true
7. }
```

- Status Codes

- 201 Created – Resource created.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

**GET** /api/osd/{svc\_id}/smart

- Parameters

- svc\_id (string) –

**Example request:**

```

1. GET /api/osd/{svc_id}/smart HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## Orchestrator

**GET** /api/orchestrator/status

## Display Orchestrator Status

### Example request:

```
1. GET /api/orchestrator/status HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## OsdPerfCounter

**GET**    `/api/perf_counters/osd/{service_id}`

- Parameters
  - **service\_id** (*string*) –

### Example request:

```
1. GET /api/perf_counters/osd/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## PerfCounters

GET /api/perf\_counters

## Display Perf Counters

### Example request:

```
1. GET /api/perf_counters HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## Pool

GET /api/pool

## Display Pool List

- Query Parameters
  - attrs (string) – Pool Attributes
  - stats (boolean) – Pool Stats

### Example request:

```
1. GET /api/pool HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.

- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/pool

#### Example request:

```

1. POST /api/pool HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "application_metadata": "string",
7.     "configuration": "string",
8.     "erasure_code_profile": "string",
9.     "flags": "string",
10.    "pg_num": 1,
11.    "pool": "string",
12.    "pool_type": "string",
13.    "rule_name": "string"
14. }
```

#### • Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/pool/{pool\_name}

#### • Parameters

- **pool\_name (string)** –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/pool/{pool\_name}

- Parameters

- **pool\_name (string)** –

Query Parameters

- **attrs (string)** –
- **stats (boolean)** –

**Example request:**

```
1. GET /api/pool/{pool_name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/pool/{pool\_name}

- Parameters

- **pool\_name (string)** –

**Example request:**

```
1. PUT /api/pool/{pool_name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "application_metadata": "string",
```

```

7.      "configuration": "string",
8.      "flags": "string"
9.  }

```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/pool/{pool\_name}/configuration

- Parameters

- **pool\_name (string)** –

**Example request:**

```

1. GET /api/pool/{pool_name}/configuration HTTP/1.1
2. Host: example.com

```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

# Prometheus

GET /api/prometheus

## Example request:

```
1. GET /api/prometheus HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

GET /api/prometheus/rules

## Example request:

```
1. GET /api/prometheus/rules HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

POST /api/prometheus/silence

- Status Codes
  - **201 Created** – Resource created.

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**DELETE** [/api/prometheus/silence/{s\\_id}](/api/prometheus/silence/{s_id})

- Parameters

- **s\_id** (*string*) –

#### Status Codes

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [204 No Content](#) – Resource deleted.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET** </api/prometheus/silences>

#### Example request:

```
1. GET /api/prometheus/silences HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.

- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## PrometheusNotifications

GET /api/prometheus/notifications

### Example request:

- ```
1. GET /api/prometheus/notifications HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Rbd

GET /api/block/image

### Display Rbd Images

- Query Parameters

- **pool\_name (string)** – Pool Name

### Example request:

- ```
1. GET /api/block/image HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/block/image

#### Example request:

```

1. POST /api/block/image HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "configuration": "string",
7.     "data_pool": "string",
8.     "features": "string",
9.     "name": "string",
10.    "namespace": "string",
11.    "obj_size": 1,
12.    "pool_name": "string",
13.    "size": 1,
14.    "stripe_count": 1,
15.    "stripe_unit": "string"
16. }
```

#### • Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/block/image/clone\_format\_version

Return the RBD clone format version.

#### Example request:

```

1. GET /api/block/image/clone_format_version HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**      `/api/block/image/default_features`

**Example request:**

```
1. GET /api/block/image/default_features HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**      `/api/block/image/{image_spec}`

- Parameters

- **image\_spec (string)** –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/block/image/{image\_spec}

- Parameters

- **image\_spec (string)** –

**Example request:**

```
1. GET /api/block/image/{image_spec} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/block/image/{image\_spec}

- Parameters

- **image\_spec (string)** –

**Example request:**

```
1. PUT /api/block/image/{image_spec} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "configuration": "string",
7.   "features": "string",
8.   "name": "string",
9.   "size": 1
10. }
```

- Status Codes

- **200 OK** – Resource updated.

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/block/image/{image\_spec}/copy

- Parameters

- **image\_spec (string)** –

**Example request:**

```

1. POST /api/block/image/{image_spec}/copy HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "configuration": "string",
7.   "data_pool": "string",
8.   "dest_image_name": "string",
9.   "dest_namespace": "string",
10.  "dest_pool_name": "string",
11.  "features": "string",
12.  "obj_size": 1,
13.  "snapshot_name": "string",
14.  "stripe_count": 1,
15.  "stripe_unit": "string"
16. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

**POST** /api/block/image/{image\_spec}/flatten

- Parameters

- **image\_spec** (*string*) –

#### Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/block/image/{image\_spec}/move\_trash

- Move an image to the trash.

Images, even ones actively in-use by clones, can be moved to the trash and deleted at a later time.

- Parameters

- **image\_spec** (*string*) –

#### Example request:

```

1. POST /api/block/image/{image_spec}/move_trash HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "delay": 1
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdMirroring

**GET** /api/block/mirroring/site\_name

### Display Rbd Mirroring sitename

#### Example request:

```
1. GET /api/block/mirroring/site_name HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/block/mirroring/site\_name

#### Example request:

```
1. PUT /api/block/mirroring/site_name HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "site_name": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdMirroringPoolBootstrap

**POST** /api/block/mirroring/pool/{pool\_name}/bootstrap/peer

- Parameters

- **pool\_name (string)** –

**Example request:**

```

1. POST /api/block/mirroring/pool/{pool_name}/bootstrap/peer HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "direction": "string",
7.   "token": "string"
8. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/block/mirroring/pool/{pool\_name}/bootstrap/token

- Parameters

- **pool\_name (string)** –

**Status Codes**

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## RbdMirroringPoolMode

`GET /api/block/mirroring/pool/{pool_name}`

### Display Rbd Mirroring Summary

- Parameters

- **pool\_name (string)** – Pool Name

### Example request:

```
1. GET /api/block/mirroring/pool/{pool_name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

`PUT /api/block/mirroring/pool/{pool_name}`

- Parameters

- **pool\_name (string)** –

### Example request:

```

1. PUT /api/block/mirroring/pool/{pool_name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "mirror_mode": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdMirroringPoolPeer

GET /api/block/mirroring/pool/{pool\_name}/peer

- Parameters

- **pool\_name (string)** –

**Example request:**

```

1. GET /api/block/mirroring/pool/{pool_name}/peer HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/block/mirroring/pool/{pool_name}/peer`

- Parameters

- **pool\_name** (*string*) –

**Example request:**

```

1. POST /api/block/mirroring/pool/{pool_name}/peer HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "client_id": "string",
7.     "cluster_name": "string",
8.     "key": "string",
9.     "mon_host": "string"
10. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**    `/api/block/mirroring/pool/{pool_name}/peer/{peer_uuid}`

- Parameters

- **pool\_name** (*string*) –
- **peer\_uuid** (*string*) –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/block/mirroring/pool/{pool\_name}/peer/{peer\_uuid}

- Parameters

- **pool\_name** (*string*) –
- **peer\_uuid** (*string*) –

**Example request:**

```
1. GET /api/block/mirroring/pool/{pool_name}/peer/{peer_uuid} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/block/mirroring/pool/{pool\_name}/peer/{peer\_uuid}

- Parameters

- **pool\_name** (*string*) –
- **peer\_uuid** (*string*) –

**Example request:**

```
1. PUT /api/block/mirroring/pool/{pool_name}/peer/{peer_uuid} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "client_id": "string",
7.     "cluster_name": "string",
8.     "key": "string",
9.     "mon_host": "string"
```

10. }

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdMirroringSummary

GET /api/block/mirroring/summary

### Display Rbd Mirroring Summary

#### Example request:

```
1. GET /api/block/mirroring/summary HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdNamespace

GET /api/block/pool/{pool\_name}/namespace

- Parameters

- **pool\_name (string)** –

**Example request:**

```
1. GET /api/block/pool/{pool_name}/namespace HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/block/pool/{pool_name}/namespace`

- Parameters
  - **pool\_name (string)** –

**Example request:**

```
1. POST /api/block/pool/{pool_name}/namespace HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "namespace": "string"
7. }
```

- Status Codes
  - **201 Created** – Resource created.
  - **202 Accepted** – Operation is still executing. Please check the task queue.
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/block/pool/{pool\_name}/namespace/{namespace}

- Parameters

- **pool\_name** (*string*) –
- **namespace** (*string*) –

## Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdSnapshot

**POST** /api/block/image/{image\_spec}/snap

- Parameters

- **image\_spec** (*string*) –

## Example request:

```

1. POST /api/block/image/{image_spec}/snap HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "snapshot_name": "string"
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** `/api/block/image/{image_spec}/snap/{snapshot_name}`

- Parameters

- **image\_spec** (*string*) –
- **snapshot\_name** (*string*) –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** `/api/block/image/{image_spec}/snap/{snapshot_name}`

- Parameters

- **image\_spec** (*string*) –
- **snapshot\_name** (*string*) –

#### Example request:

```

1. PUT /api/block/image/{image_spec}/snap/{snapshot_name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "is_protected": true,
7.   "new_snap_name": "string"
8. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/block/image/{image_spec}/snap/{snapshot_name}/clone`

Clones a snapshot to an image

- Parameters

- **image\_spec** (*string*) –
- **snapshot\_name** (*string*) –

**Example request:**

```

1. POST /api/block/image/{image_spec}/snap/{snapshot_name}/clone HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "child_image_name": "string",
7.     "child_namespace": "string",
8.     "child_pool_name": "string",
9.     "configuration": "string",
10.    "data_pool": "string",
11.    "features": "string",
12.    "obj_size": 1,
13.    "stripe_count": 1,
14.    "stripe_unit": "string"
15. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

**POST** /api/block/image/{image\_spec}/snap/{snapshot\_name}/rollback

- Parameters

- **image\_spec** (*string*) –
- **snapshot\_name** (*string*) –

#### Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RbdTrash

**GET** /api/block/image/trash

#### Get RBD Trash Details by pool name

List all entries from trash.

- Query Parameters

- **pool\_name** (*string*) – Name of the pool

#### Example request:

- ```
1. GET /api/block/image/trash HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/block/image/trash/purge

Remove all expired images from trash.

- Query Parameters

- **pool\_name** (*string*) –

**Example request:**

```

1. POST /api/block/image/trash/purge HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "pool_name": "string"
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/block/image/trash/{image\_id\_spec}

- Delete an image from trash.

If image deferment time has not expired you can not removed it unless use force. But an actively in-use by clones or has snapshots can not be removed.

- Parameters

- **image\_id\_spec** (*string*) –

Query Parameters

- **force** (boolean) -

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/block/image/trash/{image_id_spec}/restore`

Restore an image from trash.

- Parameters

- **image\_id\_spec** (string) -

**Example request:**

```

1. POST /api/block/image/trash/{image_id_spec}/restore HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "new_image_name": "string"
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Rgw

GET /api/rgw/status

### Display RGW Status

#### Example request:

```
1. GET /api/rgw/status HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## RgwBucket

GET /api/rgw/bucket

- Query Parameters
  - stats (boolean) –

#### Example request:

```
1. GET /api/rgw/bucket HTTP/1.1  
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body

for the stack trace.

**POST** /api/rgw/bucket

### Example request:

```

1. POST /api/rgw/bucket HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "bucket": "string",
7.     "lock_enabled": "string",
8.     "lock_mode": "string",
9.     "lock_retention_period_days": "string",
10.    "lock_retention_period_years": "string",
11.    "placement_target": "string",
12.    "uid": "string",
13.    "zonegroup": "string"
14. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/rgw/bucket/{bucket}

- Parameters

- **bucket (string)** –

#### Query Parameters

- **purge\_objects (string)** –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/rgw/bucket/{bucket}

- Parameters

- **bucket (string)** –

**Example request:**

```
1. GET /api/rgw/bucket/{bucket} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/rgw/bucket/{bucket}

- Parameters

- **bucket (string)** –

**Example request:**

```
1. PUT /api/rgw/bucket/{bucket} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "bucket_id": "string",
7.   "lock_mode": "string",
8.   "lock_retention_period_days": "string",
9.   "lock_retention_period_years": "string",
```

```

10.    "mfa_delete": "string",
11.    "mfa_token_pin": "string",
12.    "mfa_token_serial": "string",
13.    "uid": "string",
14.    "versioning_state": "string"
15. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RgwDaemon

**GET**    </api/rgw/daemon>

### Display RGW Daemons

#### Example request:

```

1. GET /api/rgw/daemon HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    [/api/rgw/daemon/{svc\\_id}](/api/rgw/daemon/{svc_id})

- Parameters

- **svc\_id** (*string*) –

**Example request:**

```
1. GET /api/rgw/daemon/{svc_id} HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RgwMirrorPerfCounter

GET /api/perf\_counters/rbd-mirror/{service\_id}

- Parameters

- **service\_id** (*string*) –

**Example request:**

```
1. GET /api/perf_counters/rbd-mirror/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RgwPerfCounter

GET /api/perf\_counters/rgw/{service\_id}

- Parameters

- **service\_id** (*string*) –

**Example request:**

```
1. GET /api/perf_counters/rgw/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.

- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RgwSite

GET /api/rgw/site

- Query Parameters
  - **query** (*string*) –

### Example request:

```
1. GET /api/rgw/site HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.
  - **401 Unauthorized** – Unauthenticated access. Please login first.
  - **403 Forbidden** – Unauthorized access. Please check your permissions.
  - **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## RgwUser

GET /api/rgw/user

### Display RGW Users

### Example request:

```
1. GET /api/rgw/user HTTP/1.1
2. Host: example.com
```

- Status Codes
  - **200 OK** – OK
  - **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST** /api/rgw/user

#### Example request:

```

1. POST /api/rgw/user HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "access_key": "string",
7.     "display_name": "string",
8.     "email": "string",
9.     "generate_key": "string",
10.    "max_buckets": "string",
11.    "secret_key": "string",
12.    "suspended": "string",
13.    "uid": "string"
14. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/rgw/user/get\_emails

#### Example request:

```

1. GET /api/rgw/user/get_emails HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**DELETE**`/api/rgw/user/{uid}`

- Parameters

- **uid** (*string*) –

## Status Codes

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [204 No Content](#) – Resource deleted.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET**`/api/rgw/user/{uid}`

- Parameters

- **uid** (*string*) –

**Example request:**

1. GET `/api/rgw/user/{uid}` HTTP/1.1
2. Host: example.com

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/rgw/user/{uid}

- Parameters

- **uid** (*string*) –

**Example request:**

```

1. PUT /api/rgw/user/{uid} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "display_name": "string",
7.   "email": "string",
8.   "max_buckets": "string",
9.   "suspended": "string"
10. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE** /api/rgw/user/{uid}/capability

- Parameters

- **uid** (*string*) –

Query Parameters

- **type** (*string*) – (Required)
- **perm** (*string*) – (Required)

## Status Codes

- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [204 No Content](#) – Resource deleted.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**POST** [/api/rgw/user/{uid}/capability](#)

- Parameters

- **uid** (*string*) –

### Example request:

```

1. POST /api/rgw/user/{uid}/capability HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "perm": "string",
7.     "type": "string"
8. }
```

- Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**DELETE** [/api/rgw/user/{uid}/key](#)

- Parameters

- **uid** (*string*) -

#### Query Parameters

- **key\_type** (*string*) -
- **subuser** (*string*) -
- **access\_key** (*string*) -

#### Status Codes

- **202 Accepted** - Operation is still executing. Please check the task queue.
- **204 No Content** - Resource deleted.
- **400 Bad Request** - Operation exception. Please check the response body for details.
- **401 Unauthorized** - Unauthenticated access. Please login first.
- **403 Forbidden** - Unauthorized access. Please check your permissions.
- **500 Internal Server Error** - Unexpected error. Please check the response body for the stack trace.

POST

/api/rgw/user/{uid}/key

- Parameters

- **uid** (*string*) -

#### Example request:

```

1. POST /api/rgw/user/{uid}/key HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "access_key": "string",
7.   "generate_key": "string",
8.   "key_type": "string",
9.   "secret_key": "string",
10.  "subuser": "string"
11. }
```

- Status Codes

- **201 Created** - Resource created.
- **202 Accepted** - Operation is still executing. Please check the task queue.
- **400 Bad Request** - Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/rgw/user/{uid}/quota

- Parameters

- **uid (string)** –

**Example request:**

```
1. GET /api/rgw/user/{uid}/quota HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/rgw/user/{uid}/quota

- Parameters

- **uid (string)** –

**Example request:**

```
1. PUT /api/rgw/user/{uid}/quota HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "enabled": "string",
7.     "max_objects": "string",
8.     "max_size_kb": 1,
9.     "quota_type": "string"
10. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    `/api/rgw/user/{uid}/subuser`

- Parameters

- **uid** (*string*) –

**Example request:**

```

1. POST /api/rgw/user/{uid}/subuser HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "access": "string",
7.   "access_key": "string",
8.   "generate_secret": "string",
9.   "key_type": "string",
10.  "secret_key": "string",
11.  "subuser": "string"
12. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body

for the stack trace.

**DELETE** /api/rgw/user/{uid}/subuser/{subuser}

- param `purge_keys`

Set to False to do not purge the keys. Note, this only works for s3 subusers.

- Parameters

- **uid** (*string*) –
- **subuser** (*string*) –

Query Parameters

- **purge\_keys** (*string*) –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Role

**GET** /api/role

### Display Role list

#### Example request:

```
1. GET /api/role HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST****/api/role**

#### Example request:

```

1. POST /api/role HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "description": "string",
7.   "name": "string",
8.   "scopes_permissions": "string"
9. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE****/api/role/{name}**

- Parameters

- **name (string)** –

#### Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/role/{name}

- Parameters

- **name (string)** –

**Example request:**

```
1. GET /api/role/{name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/role/{name}

- Parameters

- **name (string)** –

**Example request:**

```
1. PUT /api/role/{name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "description": "string",
7.     "scopes_permissions": "string"
8. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**`/api/role/{name}/clone`

- Parameters

- **name** (*string*) –

**Example request:**

```

1. POST /api/role/{name}/clone HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "new_name": "string"
7. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Service

**GET**`/api/service`

- Query Parameters

- **service\_name** (*string*) –

### Example request:

```
1. GET /api/service HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**POST**    **/api/service**

- param service\_spec

The service specification as JSON.

param service\_name

The service name, e.g. 'alertmanager'.

return

None

### Example request:

```
1. POST /api/service HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "service_name": "string",
7.   "service_spec": "string"
8. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.

- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**GET**    [/api/service/known\\_types](/api/service/known_types)

Get a list of known service types, e.g. ‘alertmanager’, ‘node-exporter’, ‘osd’ or ‘rgw’.

#### Example request:

1. GET [/api/service/known\\_types](/api/service/known_types) HTTP/1.1
2. Host: example.com

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

**DELETE**    [/api/service/{service\\_name}](/api/service/{service_name})

- param service\_name

The service name, e.g. ‘mds’ or ‘crash.foo’.

return

None

- Parameters

- **service\_name (string)** –

#### Status Codes

- [202 Accepted](#) – Operation is still executing. Please check the task queue.

- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/service/{service\_name}

- Parameters

- **service\_name (string)** –

**Example request:**

```
1. GET /api/service/{service_name} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET** /api/service/{service\_name}/daemons

- Parameters

- **service\_name (string)** –

**Example request:**

```
1. GET /api/service/{service_name}/daemons HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK

- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Settings

**GET** /api/settings

### Display Settings Information

```
Get the list of available options. :param names: A comma separated list of option names that should be processed. Defaults to None . :type names: None|str :return: A list of available options. :rtype: list[dict]
```

- Query Parameters

- **names (string)** – Name of Settings

#### Example request:

```
1. GET /api/settings HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/settings

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for

details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**`/api/settings/{name}`

- Parameters

- **name (string)** –

Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**`/api/settings/{name}`

```
Get the given option. :param name: The name of the option. :return: Returns a dict containing the name, type, default value and current value of the given option. :rtype: dict
```

- Parameters

- **name (string)** –

**Example request:**

1. GET `/api/settings/{name}` HTTP/1.1
2. Host: example.com

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.

- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT** /api/settings/{name}

- Parameters

- **name (string)** –

**Example request:**

```

1. PUT /api/settings/{name} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "value": "string"
7. }
```

- Status Codes

- **200 OK** – Resource updated.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

## Summary

**GET** /api/summary

**Display Summary**

**Example request:**

```

1. GET /api/summary HTTP/1.1
2. Host: example.com
```

- Status Codes

- [200 OK](#) – OK
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## Task

[GET](#)    [/api/task](#)

### Display Tasks

- Query Parameters
  - **name** (*string*) – Task Name

#### Example request:

```
1. GET /api/task HTTP/1.1
2. Host: example.com
```

- Status Codes
  - [200 OK](#) – OK
  - [400 Bad Request](#) – Operation exception. Please check the response body for details.
  - [401 Unauthorized](#) – Unauthenticated access. Please login first.
  - [403 Forbidden](#) – Unauthorized access. Please check your permissions.
  - [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## TcmuRunnerPerfCounter

[GET](#)    [/api/perf\\_counters/tcmu-runner/{service\\_id}](#)

- Parameters
  - **service\_id** (*string*) –

#### Example request:

```
1. GET /api/perf_counters/tcmu-runner/{service_id} HTTP/1.1
2. Host: example.com
```

- Status Codes

- 200 OK – OK
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## Telemetry

**PUT**    **/api/telemetry**

Enables or disables sending data collected by the Telemetry module. :param enable: Enable or disable sending data :type enable: bool :param license\_name: License string e.g. 'sharing-1-0' to make sure the user is aware of and accepts the license for sharing Telemetry data. :type license\_name: string

**Example request:**

```
1. PUT /api/telemetry HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "enable": true,
7.     "license_name": "string"
8. }
```

- Status Codes

- 200 OK – Resource updated.
- 202 Accepted – Operation is still executing. Please check the task queue.
- 400 Bad Request – Operation exception. Please check the response body for details.
- 401 Unauthorized – Unauthenticated access. Please login first.
- 403 Forbidden – Unauthorized access. Please check your permissions.
- 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

GET /api/telemetry/report

## Get Detailed Telemetry report

Get Ceph and device report data :return: Ceph and device report data :rtype: dict

### Example request:

```
1. GET /api/telemetry/report HTTP/1.1
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

## User

GET /api/user

### Get List Of Users

### Example request:

```
1. GET /api/user HTTP/1.1
2. Host: example.com
```

- Status Codes
  - 200 OK – OK
  - 400 Bad Request – Operation exception. Please check the response body for details.
  - 401 Unauthorized – Unauthenticated access. Please login first.
  - 403 Forbidden – Unauthorized access. Please check your permissions.
  - 500 Internal Server Error – Unexpected error. Please check the response body for the stack trace.

POST /api/user

**Example request:**

```

1. POST /api/user HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "email": "string",
7.   "enabled": true,
8.   "name": "string",
9.   "password": "string",
10.  "pwdExpirationDate": "string",
11.  "pwdUpdateRequired": true,
12.  "roles": "string",
13.  "username": "string"
14. }
```

- Status Codes

- **201 Created** – Resource created.
- **202 Accepted** – Operation is still executing. Please check the task queue.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**DELETE**

/api/user/{username}

- Parameters

- **username (string)** –

## Status Codes

- **202 Accepted** – Operation is still executing. Please check the task queue.
- **204 No Content** – Resource deleted.
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.

- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**GET**    `/api/user/{username}`

- Parameters

- **username (string)** –

**Example request:**

```
1. GET /api/user/{username} HTTP/1.1
2. Host: example.com
```

- Status Codes

- **200 OK** – OK
- **400 Bad Request** – Operation exception. Please check the response body for details.
- **401 Unauthorized** – Unauthenticated access. Please login first.
- **403 Forbidden** – Unauthorized access. Please check your permissions.
- **500 Internal Server Error** – Unexpected error. Please check the response body for the stack trace.

**PUT**    `/api/user/{username}`

- Parameters

- **username (string)** –

**Example request:**

```
1. PUT /api/user/{username} HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "email": "string",
7.     "enabled": "string",
8.     "name": "string",
9.     "password": "string",
10.    "pwdExpirationDate": "string",
11.    "pwdUpdateRequired": true,
12.    "roles": "string"
13. }
```

- Status Codes

- [200 OK](#) – Resource updated.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## UserChangePassword

[POST](#) [/api/user/{username}/change\\_password](/api/user/{username}/change_password)

- Parameters

- **username** (*string*) –

**Example request:**

```

1. POST /api/user/{username}/change_password HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.   "new_password": "string",
7.   "old_password": "string"
8. }
```

- Status Codes

- [201 Created](#) – Resource created.
- [202 Accepted](#) – Operation is still executing. Please check the task queue.
- [400 Bad Request](#) – Operation exception. Please check the response body for details.
- [401 Unauthorized](#) – Unauthenticated access. Please login first.
- [403 Forbidden](#) – Unauthorized access. Please check your permissions.
- [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

## UserPasswordPolicy

POST /api/user/validate\_password

Check if the password meets the password policy. :param password: The password to validate. :param username: The name of the user (optional). :param old\_password: The old password (optional). :return: An object with properties valid, credits and valuation. 'credits' contains the password complexity credits and 'valuation' the textual summary of the validation.

### Example request:

```
1. POST /api/user/validate_password HTTP/1.1
2. Host: example.com
3. Content-Type: application/json
4.
5. {
6.     "old_password": "string",
7.     "password": "string",
8.     "username": "string"
9. }
```

- Status Codes
  - [201 Created](#) – Resource created.
  - [202 Accepted](#) – Operation is still executing. Please check the task queue.
  - [400 Bad Request](#) – Operation exception. Please check the response body for details.
  - [401 Unauthorized](#) – Unauthenticated access. Please login first.
  - [403 Forbidden](#) – Unauthorized access. Please check your permissions.
  - [500 Internal Server Error](#) – Unexpected error. Please check the response body for the stack trace.

# Alerts module

The alerts module can send simple alert messages about cluster health via e-mail. In the future, it will support other notification methods as well.

## note

This module is *not* intended to be a robust monitoring solution. The fact that it is run as part of the Ceph cluster itself is fundamentally limiting in that a failure of the ceph-mgr daemon prevents alerts from being sent. This module can, however, be useful for standalone clusters that exist in environments where existing monitoring infrastructure does not exist.

## Enabling

The `alerts` module is enabled with:

```
1. ceph mgr module enable alerts
```

## Configuration

To configure SMTP, all of the following config options must be set:

```
1. ceph config set mgr mgr/alerts/smtp_host *<smtp-server>*
2. ceph config set mgr mgr/alerts/smtp_destination *<email-address-to-send-to>*
3. ceph config set mgr mgr/alerts/smtp_sender *<from-email-address>*
```

By default, the module will use SSL and port 465. To change that,:

```
1. ceph config set mgr mgr/alerts/smtp_ssl false # if not SSL
2. ceph config set mgr mgr/alerts/smtp_port *<port-number>* # if not 465
```

To authenticate to the SMTP server, you must set the user and password:

```
1. ceph config set mgr mgr/alerts/smtp_user *<username>*
2. ceph config set mgr mgr/alerts/smtp_password *<password>*
```

By default, the name in the `From:` line is simply `Ceph`. To change that (e.g., to identify which cluster this is),:

```
1. ceph config set mgr mgr/alerts/smtp_from_name 'Ceph Cluster Foo'
```

By default, the module will check the cluster health once per minute and, if there is

a change, send a message. To change that frequency, :

```
1. ceph config set mgr mgr/alerts/interval *<interval>* # e.g., "5m" for 5 minutes
```

## Commands

To force an alert to be send immediately, :

```
1. ceph alerts send
```

# Diskprediction Module

The *diskprediction* module leverages Ceph device health check to collect disk health metrics and uses internal predictor module to produce the disk failure prediction and returns back to Ceph. It doesn't require any external server for data analysis and output results. Its internal predictor's accuracy is around 70%.

## Enabling

Run the following command to enable the *diskprediction\_local* module in the Ceph environment:

```
1. ceph mgr module enable diskprediction_local
```

To enable the local predictor:

```
1. ceph config set global device_failure_prediction_mode local
```

To disable prediction, :

```
1. ceph config set global device_failure_prediction_mode none
```

*diskprediction\_local* requires at least six datasets of device health metrics to make prediction of the devices' life expectancy. And these health metrics are collected only if health monitoring is [enabled](#).

Run the following command to retrieve the life expectancy of given device.

```
1. ceph device predict-life-expectancy <device id>
```

## Configuration

The module performs the prediction on a daily basis by default. You can adjust this interval with:

```
1. ceph config set mgr mgr/diskprediction_local/predict_interval <interval-in-seconds>
```

## Debugging

If you want to debug the DiskPrediction module mapping to Ceph logging level, use the following command.

```
1. [mgr]  
2.  
3. debug mgr = 20
```

With logging set to debug for the manager the module will print out logging message with prefix `mgr[diskprediction]` for easy filtering.

# Local Pool Module

The `localpool` module can automatically create RADOS pools that are localized to a subset of the overall cluster. For example, by default, it will create a pool for each distinct rack in the cluster. This can be useful for some deployments that want to distribute some data locally as well as globally across the cluster .

## Enabling

The `localpool` module is enabled with:

```
1. ceph mgr module enable localpool
```

## Configuring

The `localpool` module understands the following options:

- **subtree** (default: `rack`): which CRUSH subtree type the module should create a pool for.
- **failure\_domain** (default: `host`): what failure domain we should separate data replicas across.
- **pg\_num** (default: 128): number of PGs to create for each pool
- **num\_rep** (default: 3): number of replicas for each pool. (Currently, pools are always replicated.)
- **min\_size** (default: `none`): value to set `min_size` to (unchanged from Ceph's default if this option is not set)
- **prefix** (default: `by-$subtreetype-`): prefix for the pool name.

These options are set via the config-key interface. For example, to change the replication level to 2x with only 64 PGs,

```
1. ceph config set mgr mgr/localpool/num_rep 2
2. ceph config set mgr mgr/localpool/pg_num 64
```

# Restful Module

RESTful module offers the REST API access to the status of the cluster over an SSL-secured connection.

## Enabling

The `restful` module is enabled with:

```
1. ceph mgr module enable restful
```

You will also need to configure an SSL certificate below before the API endpoint is available. By default the module will accept HTTPS requests on port `8003` on all IPv4 and IPv6 addresses on the host.

## Securing

All connections to `restful` are secured with SSL. You can generate a self-signed certificate with the command:

```
1. ceph restful create-self-signed-cert
```

Note that with a self-signed certificate most clients will need a flag to allow a connection and/or suppress warning messages. For example, if the `ceph-mgr` daemon is on the same host,:

```
1. curl -k https://localhost:8003/
```

To properly secure a deployment, a certificate that is signed by the organization's certificate authority should be used. For example, a key pair can be generated with a command similar to:

```
1. openssl req -new -nodes -x509 \
2.   -subj "/O=IT/CN=ceph-mgr-restful" \
3.   -days 3650 -keyout restful.key -out restful.crt -extensions v3_ca
```

The `restful.crt` should then be signed by your organization's CA (certificate authority). Once that is done, you can set it with:

```
1. ceph config-key set mgr/restful/$name/crt -i restful.crt
2. ceph config-key set mgr/restful/$name/key -i restful.key
```

where `$name` is the name of the `ceph-mgr` instance (usually the hostname). If all

manager instances are to share the same certificate, you can leave off the `$name` portion:

1. `ceph config-key set mgr/restful/crt -i restful.crt`
2. `ceph config-key set mgr/restful/key -i restful.key`

## Configuring IP and port

Like any other RESTful API endpoint, `restful` binds to an IP and port. By default, the currently active `ceph-mgr` daemon will bind to port 8003 and any available IPv4 or IPv6 address on the host.

Since each `ceph-mgr` hosts its own instance of `restful`, it may also be necessary to configure them separately. The IP and port can be changed via the configuration key facility:

1. `ceph config set mgr mgr/restful/$name/server_addr $IP`
2. `ceph config set mgr mgr/restful/$name/server_port $PORT`

where `$name` is the ID of the `ceph-mgr` daemon (usually the hostname).

These settings can also be configured cluster-wide and not manager specific. For example, :

1. `ceph config set mgr mgr/restful/server_addr $IP`
2. `ceph config set mgr mgr/restful/server_port $PORT`

If the port is not configured, `restful` will bind to port `8003`. If the address is not configured, the `restful` will bind to `::`, which corresponds to all available IPv4 and IPv6 addresses.

## Creating an API User

To create an API user, please run the following command:

1. `ceph restful create-key <username>`

Replace `<username>` with the desired name of the user. For example, to create a user named `api` :

1. `$ ceph restful create-key api`
2. `52dff92-a103-4a10-bfce-5b60f48f764e`

The UUID generated from `ceph restful create-key api` acts as the key for the user.

To list all of your API keys, please run the following command:

```
1. ceph restful list-keys
```

The `ceph restful list-keys` command will output in JSON:

```
1. {
2.     "api": "52dff92-a103-4a10-bfce-5b60f48f764e"
3. }
```

You can use `curl` in order to test your user with the API. Here is an example:

```
1. curl -k https://api:52dff92-a103-4a10-bfce-5b60f48f764e@<ceph-mgr>:<port>/server
```

In the case above, we are using `GET` to fetch information from the `server` endpoint.

## Load balancer

Please note that `restful` will *only* start on the manager which is active at that moment. Query the Ceph cluster status to see which manager is active (e.g., `ceph mgr dump`). In order to make the API available via a consistent URL regardless of which manager daemon is currently active, you may want to set up a load balancer front-end to direct traffic to whichever manager endpoint is available.

## Available methods

You can navigate to the `/doc` endpoint for full list of available endpoints and HTTP methods implemented for each endpoint.

For example, if you want to use the PATCH method of the `/osd/<id>` endpoint to set the state `up` of the OSD id `1`, you can use the following curl command:

```
echo -En '{"up": true}' | curl --request PATCH --data @- --silent --insecure --user <user> 'https://<ceph-mgr>:<port>/osd/1'
```

or you can use python to do so:

```
1. $ python
2. >>> import requests
3. >>> result = requests.patch(
4.         'https://<ceph-mgr>:<port>/osd/1',
5.         json={"up": True},
6.         auth=("<user>", "<password>")
7.     )
8. >>> print result.json()
```

Some of the other endpoints implemented in the `restful` module include

- `/config/cluster` : **GET**
- `/config/osd` : **GET, PATCH**
- `/crush/rule` : **GET**
- `/mon` : **GET**
- `/osd` : **GET**
- `/pool` : **GET, POST**
- `/pool/<arg>` : **DELETE, GET, PATCH**
- `/request` : **DELETE, GET, POST**
- `/request/<arg>` : **DELETE, GET**
- `/server` : **GET**

## The `/request` endpoint

You can use the `/request` endpoint to poll the state of a request you scheduled with any **DELETE**, **POST** or **PATCH** method. These methods are by default asynchronous since it may take longer for them to finish execution. You can modify this behaviour by appending `?wait=1` to the request url. The returned request will then always be completed.

The **POST** method of the `/request` method provides a passthrough for the ceph mon commands as defined in `src/mon/MonCommands.h`. Let's consider the following command:

```
1. COMMAND("osd ls" \
2.          "name=epoch,type=CephInt,range=0,req=false", \
3.          "show all OSD ids", "osd", "r", "cli,rest")
```

The **prefix** is **osd ls**. The optional argument's name is **epoch** and it is of type `CephInt`, i.e. `integer`. This means that you need to do the following **POST** request to schedule the command:

```
1. $ python
2. >> import requests
3. >> result = requests.post(
4.         'https://<ceph-mgr>:<port>/request',
5.         json={'prefix': 'osd ls', 'epoch': 0},
6.         auth=("<user>", "<password>")
7.     )
8. >> print result.json()
```

# Zabbix Module

The Zabbix module actively sends information to a Zabbix server like:

- Ceph status
- I/O operations
- I/O bandwidth
- OSD status
- Storage utilization

## Requirements

The module requires that the `zabbix_sender` executable is present on *all* machines running `ceph-mgr`. It can be installed on most distributions using the package manager.

## Dependencies

Installing `zabbix_sender` can be done under Ubuntu or CentOS using either `apt` or `dnf`.

On Ubuntu Xenial:

```
1. apt install zabbix-agent
```

On Fedora:

```
1. dnf install zabbix-sender
```

## Enabling

You can enable the `zabbix` module with:

```
1. ceph mgr module enable zabbix
```

## Configuration

Two configuration keys are vital for the module to work:

- `zabbix_host`
- `identifier` (optional)

The parameter `zabbix_host` controls the hostname of the Zabbix server to which `zabbix_sender` will send the items. This can be a IP-Address if required by your installation.

The `identifier` parameter controls the identifier/hostname to use as source when sending items to Zabbix. This should match the name of the `Host` in your Zabbix server.

When the `identifier` parameter is not configured the `ceph-<fsid>` of the cluster will be used when sending data to Zabbix.

This would for example be `ceph-c4d32a99-9e80-490f-bd3a-1d22d8a7d354`

Additional configuration keys which can be configured and their default values:

- `zabbix_port`: 10051
- `zabbix_sender`: `/usr/bin/zabbix_sender`
- `interval`: 60
- `discovery_interval`: 100

## Configuration keys

Configuration keys can be set on any machine with the proper cephx credentials, these are usually Monitors where the `client.admin` key is present.

```
1. ceph zabbix config-set <key> <value>
```

For example:

```
1. ceph zabbix config-set zabbix_host zabbix.localdomain
2. ceph zabbix config-set identifier ceph.eu-ams02.local
```

The current configuration of the module can also be shown:

```
1. ceph zabbix config-show
```

## Template

A `template`. (XML) to be used on the Zabbix server can be found in the source directory of the module.

This template contains all items and a few triggers. You can customize the triggers afterwards to fit your needs.

## Multiple Zabbix servers

It is possible to instruct zabbix module to send data to multiple Zabbix servers.

Parameter `zabbix_host` can be set with multiple hostnames separated by commas. Hostnames (or IP addresses) can be followed by colon and port number. If a port number is not present module will use the port number defined in `zabbix_port`.

For example:

```
1. ceph zabbix config-set zabbix_host "zabbix1,zabbix2:2222,zabbix3:3333"
```

## Manually sending data

If needed the module can be asked to send data immediately instead of waiting for the interval.

This can be done with this command:

```
1. ceph zabbix send
```

The module will now send its latest data to the Zabbix server.

Items discovery is accomplished also via `zabbix_sender`, and runs every `discovery_interval * interval` seconds. If you wish to launch discovery manually, this can be done with this command:

```
1. ceph zabbix discovery
```

## Debugging

Should you want to debug the Zabbix module increase the logging level for `ceph-mgr` and check the logs.

```
1. [mgr]  
2.     debug mgr = 20
```

With logging set to debug for the manager the module will print various logging lines prefixed with `mgr[zabbix]` for easy filtering.

# Prometheus Module

Provides a Prometheus exporter to pass on Ceph performance counters from the collection point in ceph-mgr. Ceph-mgr receives MMgrReport messages from all MgrClient processes (mons and OSDs, for instance) with performance counter schema data and actual counter data, and keeps a circular buffer of the last N samples. This module creates an HTTP endpoint (like all Prometheus exporters) and retrieves the latest sample of every counter when polled (or “scraped” in Prometheus terminology). The HTTP path and query parameters are ignored; all extant counters for all reporting entities are returned in text exposition format. (See the Prometheus [documentation](#).)

## Enabling prometheus output

The *prometheus* module is enabled with:

```
1. ceph mgr module enable prometheus
```

## Configuration

### Note

The Prometheus manager module needs to be restarted for configuration changes to be applied.

By default the module will accept HTTP requests on port `9283` on all IPv4 and IPv6 addresses on the host. The port and listen address are both configurable with `ceph config set`, with keys `mgr/prometheus/server_addr` and `mgr/prometheus/server_port`. This port is registered with Prometheus’s [registry](#).

```
1. ceph config set mgr mgr/prometheus/server_addr 0.0.0.0
2. ceph config set mgr mgr/prometheus/server_port 9283
```

### Warning

The `scrape_interval` of this module should always be set to match Prometheus’ scrape interval to work properly and not cause any issues.

The Prometheus manager module is, by default, configured with a scrape interval of 15 seconds. The scrape interval in the module is used for caching purposes and to determine when a cache is stale.

It is not recommended to use a scrape interval below 10 seconds. It is recommended to use 15 seconds as scrape interval, though, in some cases it might be useful to increase the scrape interval.

To set a different scrape interval in the Prometheus module, set `scrape_interval` to the desired value:

```
1. ceph config set mgr mgr/prometheus/scrape_interval 20
```

On large clusters (>1000 OSDs), the time to fetch the metrics may become significant. Without the cache, the Prometheus manager module could, especially in conjunction with multiple Prometheus instances, overload the manager and lead to unresponsive or crashing Ceph manager instances. Hence, the cache is enabled by default and cannot be disabled. This means that there is a possibility that the cache becomes stale. The cache is considered stale when the time to fetch the metrics from Ceph exceeds the configured `scrape_interval`.

If that is the case, a warning will be logged and the module will either

- respond with a 503 HTTP status code (service unavailable) or,
- it will return the content of the cache, even though it might be stale.

This behavior can be configured. By default, it will return a 503 HTTP status code (service unavailable). You can set other options using the `ceph config set` commands.

To tell the module to respond with possibly stale data, set it to `return`:

```
1. ceph config set mgr mgr/prometheus/stale_cache_strategy return
```

To tell the module to respond with “service unavailable”, set it to `fail`:

```
1. ceph config set mgr mgr/prometheus/stale_cache_strategy fail
```

## RBD IO statistics

The module can optionally collect RBD per-image IO statistics by enabling dynamic OSD performance counters. The statistics are gathered for all images in the pools that are specified in the `mgr/prometheus/rbd_stats_pools` configuration parameter. The parameter is a comma or space separated list of `pool[/namespace]` entries. If the namespace is not specified the statistics are collected for all namespaces in the pool.

Example to activate the RBD-enabled pools `pool1`, `pool2` and `poolN`:

```
1. ceph config set mgr mgr/prometheus/rbd_stats_pools "pool1,pool2,poolN"
```

The module makes the list of all available images scanning the specified pools and namespaces and refreshes it periodically. The period is configurable via the `mgr/prometheus/rbd_stats_pools_refresh_interval` parameter (in sec) and is 300 sec (5 minutes) by default. The module will force refresh earlier if it detects statistics from a previously unknown RBD image.

Example to turn up the sync interval to 10 minutes:

```
1. ceph config set mgr mgr/prometheus/rbd_stats_pools_refresh_interval 600
```

## Statistic names and labels

The names of the stats are exactly as Ceph names them, with illegal characters `.`, `-` and `::` translated to `_`, and `ceph_` prefixed to all names.

All *daemon* statistics have a `ceph_daemon` label such as “osd.123” that identifies the type and ID of the daemon they come from. Some statistics can come from different types of daemon, so when querying e.g. an OSD’s RocksDB stats, you would probably want to filter on `ceph_daemon` starting with “osd” to avoid mixing in the monitor rocksdb stats.

The *cluster* statistics (i.e. those global to the Ceph cluster) have labels appropriate to what they report on. For example, metrics relating to pools have a `pool_id` label.

The long running averages that represent the histograms from core Ceph are represented by a pair of `<name>_sum` and `<name>_count` metrics. This is similar to how histograms are represented in Prometheus and they can also be treated [similarly](#).

## Pool and OSD metadata series

Special series are output to enable displaying and querying on certain metadata fields.

Pools have a `ceph_pool_metadata` field like this:

```
1. ceph_pool_metadata{pool_id="2",name="cephfs_metadata_a"} 1.0
```

OSDs have a `ceph_osd_metadata` field like this:

```
ceph_osd_metadata{cluster_addr="172.21.9.34:6802/19096",device_class="ssd",ceph_daemon="osd.0",public_addr="172.21.9.34:6802/19096"} 1.0
```

## Correlating drive statistics with node\_exporter

The prometheus output from Ceph is designed to be used in conjunction with the generic host monitoring from the Prometheus node\_exporter.

To enable correlation of Ceph OSD statistics with node\_exporter’s drive statistics, special series are output like this:

```
1. ceph_disk_occupation{ceph_daemon="osd.0",device="sdd", exported_instance="myhost"}
```

To use this to get disk statistics by OSD ID, use either the `and` operator or the `*` operator in your prometheus query. All metadata metrics (like `ceph_disk_occupation`) have the value 1 so they act neutral with `*`. Using `*` allows to use `group_left` and `group_right` grouping modifiers, so that the resulting metric has additional labels from one side of the query.

See the [prometheus documentation](#) for more information about constructing queries.

The goal is to run a query like

```
1. rate(node_disk_bytes_written[30s]) and on (device,instance) ceph_disk_occupation{ceph_daemon="osd.0"}
```

Out of the box the above query will not return any metrics since the `instance` labels of both metrics don't match. The `instance` label of `ceph_disk_occupation` will be the currently active MGR node.

The following two section outline two approaches to remedy this.

## Use `label_replace`

The `label_replace` function ([cp. `label\_replace` documentation](#)) can add a label to, or alter a label of, a metric within a query.

To correlate an OSD and its disks write rate, the following query can be used:

```
label_replace(rate(node_disk_bytes_written[30s]), "exported_instance", "$1", "instance", "(.*)_.*") and on
1. (device,exported_instance) ceph_disk_occupation{ceph_daemon="osd.0"}
```

## Configuring Prometheus server

### `honor_labels`

To enable Ceph to output properly-labeled data relating to any host, use the `honor_labels` setting when adding the ceph-mgr endpoints to your prometheus configuration.

This allows Ceph to export the proper `instance` label without prometheus overwriting it. Without this setting, Prometheus applies an `instance` label that includes the hostname and port of the endpoint that the series came from. Because Ceph clusters have multiple manager daemons, this results in an `instance` label that changes spuriously when the active manager daemon changes.

If this is undesirable a custom `instance` label can be set in the Prometheus target configuration: you might wish to set it to the hostname of your first mgr daemon, or something completely arbitrary like "ceph\_cluster".

## node\_exporter hostname labels

Set your `instance` labels to match what appears in Ceph's OSD metadata in the `instance` field. This is generally the short hostname of the node.

This is only necessary if you want to correlate Ceph stats with host stats, but you may find it useful to do it in all cases in case you want to do the correlation in the future.

## Example configuration

This example shows a single node configuration running ceph-mgr and node\_exporter on a server called `senta04`. Note that this requires one to add an appropriate and unique `instance` label to each `node_exporter` target.

This is just an example: there are other ways to configure prometheus scrape targets and label rewrite rules.

### `prometheus.yml`

```

1. global:
2.   scrape_interval:      15s
3.   evaluation_interval: 15s
4.
5.   scrape_configs:
6.     - job_name: 'node'
7.       file_sd_configs:
8.         - files:
9.           - node_targets.yml
10.    - job_name: 'ceph'
11.      honor_labels: true
12.      file_sd_configs:
13.        - files:
14.          - ceph_targets.yml

```

### `ceph_targets.yml`

```

1. [
2.   {
3.     "targets": [ "senta04.mydomain.com:9283" ],
4.     "labels": {}
5.   }
6. ]

```

### `node_targets.yml`

```

1. [
2.   {

```

```
3.     "targets": [ "senta04.mydomain.com:9100" ],
4.     "labels": {
5.       "instance": "senta04"
6.     }
7.   }
8. ]
```

## Notes

---

Counters and gauges are exported; currently histograms and long-running averages are not. It's possible that Ceph's 2-D histograms could be reduced to two separate 1-D histograms, and that long-running averages could be exported as Prometheus' Summary type.

Timestamps, as with many Prometheus exporters, are established by the server's scrape time (Prometheus expects that it is polling the actual counter process synchronously). It is possible to supply a timestamp along with the stat report, but the Prometheus team strongly advises against this. This means that timestamps will be delayed by an unpredictable amount; it's not clear if this will be problematic, but it's worth knowing about.

# Influx Module

The `influx` module continuously collects and sends time series data to an `influxdb` database.

The `influx` module was introduced in the 13.x *Mimic* release.

## Enabling

To enable the module, use the following command:

```
1. ceph mgr module enable influx
```

If you wish to subsequently disable the module, you can use the equivalent *disable* command:

```
1. ceph mgr module disable influx
```

## Configuration

For the `influx` module to send statistics to an InfluxDB server, it is necessary to configure the servers address and some authentication credentials.

Set configuration values using the following command:

```
1. ceph config set mgr mgr/influx/<key> <value>
```

The most important settings are `hostname`, `username` and `password`. For example, a typical configuration might look like this:

```
1. ceph config set mgr mgr/influx/hostname influx.mydomain.com
2. ceph config set mgr mgr/influx/username admin123
3. ceph config set mgr mgr/influx/password p4ssw0rd
```

Additional optional configuration settings are:

`interval`

Time between reports to InfluxDB. Default 30 seconds.

`database`

InfluxDB database name. Default “ceph”. You will need to create this database and grant write privileges to the configured username or the username must have admin

privileges to create it.

port

InfluxDB server port. Default 8086

ssl

Use https connection for InfluxDB server. Use "true" or "false". Default false

verify\_ssl

Verify https cert for InfluxDB server. Use "true" or "false". Default true

threads

How many worker threads should be spawned for sending data to InfluxDB. Default is 5

batch\_size

How big batches of data points should be when sending to InfluxDB. Default is 5000

## Debugging

---

By default, a few debugging statements as well as error statements have been set to print in the log files. Users can add more if necessary. To make use of the debugging option in the module:

- Add this to the ceph.conf file.:

```
1. [mgr]
2.   debug_mgr = 20
```

- Use this command `ceph influx self-test` .
- Check the log files. Users may find it easier to filter the log files using `mgr[influx]`.

## Interesting counters

---

The following tables describe a subset of the values output by this module.

## Pools

Counter	Description
stored	Bytes stored in the pool not including copies
max_avail	Max available number of bytes in the pool

objects	Number of objects in the pool
wr_bytes	Number of bytes written in the pool
dirty	Number of bytes dirty in the pool
rd_bytes	Number of bytes read in the pool
stored_raw	Bytes used in pool including copies made

## OSDs

Counter	Description
op_w	Client write operations
op_in_bytes	Client operations total write size
op_r	Client read operations
op_out_bytes	Client operations total read size

Counter	Description
op_wip	Replication operations currently being processed (primary)
op_latency	Latency of client operations (including queue time)
op_process_latency	Latency of client operations (excluding queue time)
op_prepare_latency	Latency of client operations (excluding queue time and wait for finished)
op_r_latency	Latency of read operation (including queue time)
op_r_process_latency	Latency of read operation (excluding queue time)
op_w_in_bytes	Client data written
op_w_latency	Latency of write operation (including queue time)
op_w_process_latency	Latency of write operation (excluding queue time)
	Latency of write operations (excluding queue time and

	wait for finished)
op_rw	Client read-modify-write operations
op_rw_in_bytes	Client read-modify-write operations write in
op_rw_out_bytes	Client read-modify-write operations read out
op_rw_latency	Latency of read-modify-write operation (including queue time)
op_rw_process_latency	Latency of read-modify-write operation (excluding queue time)
op_rw_prepare_latency	Latency of read-modify-write operations (excluding queue time and wait for finished)
op_before_queue_op_lat	Latency of IO before calling queue (before really queue into ShardedOpWq) op_before_dequeue_op_lat
op_before_dequeue_op_lat	Latency of IO before calling dequeue_op(already dequeued and get PG lock)

Latency counters are measured in microseconds unless otherwise specified in the description.

# Hello World Module

This is a simple module skeleton for documentation purposes.

## Enabling

The `hello` module is enabled with:

```
1. ceph mgr module enable hello
```

To check that it is enabled, run:

```
1. ceph mgr module ls
```

After editing the module file (found in `src/pybind/mgr/hello/module.py`), you can see changes by running:

```
1. ceph mgr module disable hello  
2. ceph mgr module enable hello
```

or:

```
1. init-ceph restart mgr
```

To execute the module, run:

```
1. ceph hello
```

The log is found at:

```
1. build/out/mgr.x.log
```

## Documenting

After adding a new mgr module, be sure to add its documentation to `doc/mgr/module_name.rst`. Also, add a link to your new module into `doc/mgr/index.rst`.

# Telegraf Module

The Telegraf module collects and sends statistics series to a Telegraf agent.

The Telegraf agent can buffer, aggregate, parse and process the data before sending it to an output which can be InfluxDB, ElasticSearch and many more.

Currently the only way to send statistics to Telegraf from this module is to use the socket listener. The module can send statistics over UDP, TCP or a UNIX socket.

The Telegraf module was introduced in the 13.x *Mimic* release.

## Enabling

To enable the module, use the following command:

```
1. ceph mgr module enable telegraf
```

If you wish to subsequently disable the module, you can use the corresponding *disable* command:

```
1. ceph mgr module disable telegraf
```

## Configuration

For the telegraf module to send statistics to a Telegraf agent it is required to configure the address to send the statistics to.

Set configuration values using the following command:

```
1. ceph telegraf config-set <key> <value>
```

The most important settings are `address` and `interval`.

For example, a typical configuration might look like this:

```
1. ceph telegraf config-set address udp://:8094
2. ceph telegraf config-set interval 10
```

The default values for these configuration keys are:

- `address`: `unixgram:///tmp/telegraf.sock`
- `interval`: `15`

# Socket Listener

The module only supports sending data to Telegraf through the socket listener of the Telegraf module using the Influx data format.

A typical Telegraf configuration might be:

```
[[inputs.socket_listener]] # service_address = "tcp://:8094" # service_address = "tcp://127.0.0.1:http" #
service_address = "tcp4://:8094" # service_address = "tcp6://:8094" # service_address = "tcp6://[2001:db8::1]:8094"
service_address = "udp://:8094" # service_address = "udp4://:8094" # service_address = "udp6://:8094" #
service_address = "unix:///tmp/telegraf.sock" # service_address = "unixgram:///tmp/telegraf.sock" data_format =
"influx"
```

In this case the address configuration option for the module would need to be set to:

```
udp://:8094
```

Refer to the Telegraf documentation for more configuration options.

# Telemetry Module

The telemetry module sends anonymous data about the cluster back to the Ceph developers to help understand how Ceph is used and what problems users may be experiencing.

This data is visualized on [public dashboards](#) that allow the community to quickly see summary statistics on how many clusters are reporting, their total capacity and OSD count, and version distribution trends.

## Channels

The telemetry report is broken down into several “channels”, each with a different type of information. Assuming telemetry has been enabled, individual channels can be turned on and off. (If telemetry is off, the per-channel setting has no effect.)

- **basic** (default: on): Basic information about the cluster

- capacity of the cluster
- number of monitors, managers, OSDs, MDSS, object gateways, or other daemons
- software version currently being used
- number and types of RADOS pools and CephFS file systems
- names of configuration options that have been changed from their default (but *not* their values)

- **crash** (default: on): Information about daemon crashes, including

- type of daemon
- version of the daemon
- operating system (OS distribution, kernel version)
- stack trace identifying where in the Ceph code the crash occurred

- **device** (default: on): Information about device metrics, including

- anonymized SMART metrics

- **ident** (default: off): User-provided identifying information about the cluster

- cluster description

- contact email address

The data being reported does *not* contain any sensitive data like pool names, object names, object contents, hostnames, or device serial numbers.

It contains counters and statistics on how the cluster has been deployed, the version of Ceph, the distribution of the hosts and other parameters which help the project to gain a better understanding of the way Ceph is used.

Data is sent secured to <https://telemetry.ceph.com>.

## Sample report

You can look at what data is reported at any time with the command:

1. ceph telemetry show

To protect your privacy, device reports are generated separately, and data such as hostname and device serial number is anonymized. The device telemetry is sent to a different endpoint and does not associate the device data with a particular cluster. To see a preview of the device report use the command:

1. ceph telemetry show-device

Please note: In order to generate the device report we use Smartmontools version 7.0 and up, which supports JSON output. If you have any concerns about privacy with regard to the information included in this report, please contact the Ceph developers.

## Channels

Individual channels can be enabled or disabled with:

1. ceph config set mgr mgr/telemetry/channel\_ident false
2. ceph config set mgr mgr/telemetry/channel\_basic false
3. ceph config set mgr mgr/telemetry/channel\_crash false
4. ceph config set mgr mgr/telemetry/channel\_device false
5. ceph telemetry show
6. ceph telemetry show-device

## Enabling Telemetry

To allow the *telemetry* module to start sharing data:

1. ceph telemetry on

Please note: Telemetry data is licensed under the Community Data License Agreement - Sharing - Version 1.0 (<https://cdla.io/sharing-1-0/>). Hence, telemetry module can be enabled only after you add '-license sharing-1-0' to the 'ceph telemetry on' command.

Telemetry can be disabled at any time with:

```
1. ceph telemetry off
```

## Interval

The module compiles and sends a new report every 24 hours by default. You can adjust this interval with:

```
1. ceph config set mgr mgr/telemetry/interval 72      # report every three days
```

## Status

To see the current configuration:

```
1. ceph telemetry status
```

## Manually sending telemetry

To ad hoc send telemetry data:

```
1. ceph telemetry send
```

In case telemetry is not enabled (with 'ceph telemetry on'), you need to add '-license sharing-1-0' to 'ceph telemetry send' command.

## Sending telemetry through a proxy

If the cluster cannot directly connect to the configured telemetry endpoint (default `telemetry.ceph.com`), you can configure a HTTP/HTTPS proxy server with:

```
1. ceph config set mgr mgr/telemetry/proxy https://10.0.0.1:8080
```

You can also include a `user:pass` if needed:

```
1. ceph config set mgr mgr/telemetry/proxy https://ceph:telemetry@10.0.0.1:8080
```

## Contact and Description

A contact and description can be added to the report. This is completely optional, and disabled by default.:

```
1. ceph config set mgr mgr/telemetry/contact 'John Doe <john.doe@example.com>'  
2. ceph config set mgr mgr/telemetry/description 'My first Ceph cluster'  
3. ceph config set mgr mgr/telemetry/channel_ident true
```

# iostat

This module shows the current throughput and IOPS done on the Ceph cluster.

## Enabling

To check if the *iostat* module is enabled, run:

```
1. ceph mgr module ls
```

The module can be enabled with:

```
1. ceph mgr module enable iostat
```

To execute the module, run:

```
1. ceph iostat
```

To change the frequency at which the statistics are printed, use the `-p` option:

```
1. ceph iostat -p <period in seconds>
```

For example, use the following command to print the statistics every 5 seconds:

```
1. ceph iostat -p 5
```

To stop the module, press Ctrl-C.

# Crash Module

The crash module collects information about daemon crashdumps and stores it in the Ceph cluster for later analysis.

Daemon crashdumps are dumped in /var/lib/ceph/crash by default; this can be configured with the option ‘crash dir’. Crash directories are named by time and date and a randomly-generated UUID, and contain a metadata file ‘meta’ and a recent log file, with a “crash\_id” that is the same. This module allows the metadata about those dumps to be persisted in the monitors’ storage.

## Enabling

The *crash* module is enabled with:

```
1. ceph mgr module enable crash
```

## Commands

```
1. ceph crash post -i <metafile>
```

Save a crash dump. The metadata file is a JSON blob stored in the crash dir as `meta`. As usual, the `ceph` command can be invoked with `-i -`, and will read from `stdin`.

```
1. ceph crash rm <crashid>
```

Remove a specific crash dump.

```
1. ceph crash ls
```

List the timestamp/uuid crashids for all new and archived crash info.

```
1. ceph crash ls-new
```

List the timestamp/uuid crashids for all newcrash info.

```
1. ceph crash stat
```

Show a summary of saved crash info grouped by age.

```
1. ceph crash info <crashid>
```

Show all details of a saved crash.

```
1. ceph crash prune <keep>
```

Remove saved crashes older than ‘keep’ days. <keep> must be an integer.

```
1. ceph crash archive <crashid>
```

Archive a crash report so that it is no longer considered for the `RECENT_CRASH` health check and does not appear in the `crash ls-new` output (it will still appear in the `crash ls` output).

```
1. ceph crash archive-all
```

Archive all new crash reports.

## Options

---

- `mgr/crash/warn_recent_interval` [default: 2 weeks] controls what constitutes “recent” for the purposes of raising the `RECENT_CRASH` health warning.
- `mgr/crash/retain_interval` [default: 1 year] controls how long crash reports are retained by the cluster before they are automatically purged.

# Insights Module

The insights module collects and exposes system information to the Insights Core data analysis framework. It is intended to replace explicit interrogation of Ceph CLIs and daemon admin sockets, reducing the API surface that Insights depends on. The insights reports contains the following:

- **Health reports.** In addition to reporting the current health of the cluster, the insights module reports a summary of the last 24 hours of health checks. This feature is important for catching cluster health issues that are transient and may not be present at the moment the report is generated. Health checks are deduplicated to avoid unbounded data growth.
- **Crash reports.** A summary of any daemon crashes in the past 24 hours is included in the insights report. Crashes are reported as the number of crashes per daemon type (e.g. ceph-osd) within the time window. Full details of a crash may be obtained using the [crash module](#).
- Software version, storage utilization, cluster maps, placement group summary, monitor status, cluster configuration, and OSD metadata.

## Enabling

The *insights* module is enabled with:

```
1. ceph mgr module enable insights
```

## Commands

```
1. ceph insights
```

Generate the full report.

```
1. ceph insights prune-health <hours>
```

Remove historical health data older than <hours>. Passing 0 for <hours> will clear all health data.

This command is useful for cleaning the health history before automated nightly reports are generated, which may contain spurious health checks accumulated while performing system maintenance, or other health checks that have been resolved. There is no need to prune health data to reclaim storage space; garbage collection is performed regularly to remove old health data from persistent storage.

# Orchestrator CLI

This module provides a command line interface (CLI) to orchestrator modules ( `ceph-mgr` modules which interface with external orchestration services).

As the orchestrator CLI unifies multiple external orchestrators, a common nomenclature for the orchestrator module is needed.

<i>host</i>	hostname (not DNS name) of the physical host. Not the podname, container name, or hostname inside the container.
<i>service type</i>	The type of the service. e.g., nfs, mds, osd, mon, rgw, mgr, iscsi
<i>service</i>	A logical service, Typically comprised of multiple service instances on multiple hosts for HA <ul style="list-style-type: none"> <li>• <code>fs_name</code> for mds type</li> <li>• <code>rgw_zone</code> for rgw type</li> <li>• <code>ganesha_cluster_id</code> for nfs type</li> </ul>
<i>daemon</i>	A single instance of a service. Usually a daemon, but maybe not (e.g., might be a kernel service like LIO or knfsd or whatever) This identifier should uniquely identify the instance

The relation between the names is the following:

- A *service* has a specific *service type*
- A *daemon* is a physical instance of a *service type*

## Note

Orchestrator modules may only implement a subset of the commands listed below. Also, the implementation of the commands may differ between modules.

## Status

### 1. `ceph orch status`

Show current orchestrator mode and high-level status (whether the orchestrator plugin is available and operational)

## Host Management

List hosts associated with the cluster:

```
1. ceph orch host ls
```

Add and remove hosts:

```
1. ceph orch host add <hostname> [<addr>] [<labels>...]
2. ceph orch host rm <hostname>
```

For cephadm, see also [Fully qualified domain names vs bare host names](#).

## Host Specification

Many hosts can be added at once using `ceph orch apply -i` by submitting a multi-document YAML file:

```
1. ---
2. service_type: host
3. addr: node-00
4. hostname: node-00
5. labels:
6. - example1
7. - example2
8. ---
9. service_type: host
10. addr: node-01
11. hostname: node-01
12. labels:
13. - grafana
14. ---
15. service_type: host
16. addr: node-02
17. hostname: node-02
```

This can be combined with service specifications (below) to create a cluster spec file to deploy a whole cluster in one command. see `cephadm bootstrap --apply-spec` also to do this during bootstrap. Cluster SSH Keys must be copied to hosts prior to adding them.

## OSD Management

---

### List Devices

Print a list of discovered devices, grouped by host and optionally filtered to a particular host:

```
1. ceph orch device ls [--host=...] [--refresh]
```

**Example:**

1.	HOST	PATH	TYPE	SIZE	DEVICE	AVAIL	REJECT REASONS
2.	master	/dev/vda	hdd	42.0G		False	locked
3.	node1	/dev/vda	hdd	42.0G		False	locked
4.	node1	/dev/vdb	hdd	8192M	387836	False	locked, LVM detected, Insufficient space (<5GB) on vgs
5.	node1	/dev/vdc	hdd	8192M	450575	False	locked, LVM detected, Insufficient space (<5GB) on vgs
6.	node3	/dev/vda	hdd	42.0G		False	locked
7.	node3	/dev/vdb	hdd	8192M	395145	False	LVM detected, locked, Insufficient space (<5GB) on vgs
8.	node3	/dev/vdc	hdd	8192M	165562	False	LVM detected, locked, Insufficient space (<5GB) on vgs
9.	node2	/dev/vda	hdd	42.0G		False	locked
10.	node2	/dev/vdb	hdd	8192M	672147	False	LVM detected, Insufficient space (<5GB) on vgs, locked
11.	node2	/dev/vdc	hdd	8192M	228094	False	LVM detected, Insufficient space (<5GB) on vgs, locked

## Erase Devices (Zap Devices)

Erase (zap) a device so that it can be reused. `zap` calls `ceph-volume zap` on the remote host.

```
1. orch device zap <hostname> <path>
```

Example command:

```
1. ceph orch device zap my_hostname /dev/sdx
```

### Note

Cephadm orchestrator will automatically deploy drives that match the DriveGroup in your OSDSpec if the unmanaged flag is unset. For example, if you use the `all-available-devices` option when creating OSDs, when you `zap` a device the cephadm orchestrator will automatically create a new OSD in the device . To disable this behavior, see [Create OSDs](#).

## Create OSDs

Create OSDs on a set of devices on a single host:

```
1. ceph orch daemon add osd <host>:device1,device2
```

Another way of doing it is using `apply` interface:

```
1. ceph orch apply osd -i <json_file/yaml_file> [--dry-run]
```

where the `json_file/yaml_file` is a DriveGroup specification. For a more in-depth guide to DriveGroups please refer to [OSD Service Specification](#)

`dry-run` will cause the orchestrator to present a preview of what will happen without actually creating the OSDs.

Example:

```
1. # ceph orch apply osd --all-available-devices --dry-run
2. NAME          HOST   DATA   DB WAL
3. all-available-devices node1 /dev/vdb - -
4. all-available-devices node2 /dev/vdc - -
5. all-available-devices node3 /dev/vdd - -
```

When the parameter `all-available-devices` or a DriveGroup specification is used, a cephadm service is created. This service guarantees that all available devices or devices included in the DriveGroup will be used for OSDs. Note that the effect of `--all-available-devices` is persistent; that is, drives which are added to the system or become available (say, by zapping) after the command is complete will be automatically found and added to the cluster.

That is, after using:

```
1. ceph orch apply osd --all-available-devices
```

- If you add new disks to the cluster they will automatically be used to create new OSDs.
- A new OSD will be created automatically if you remove an OSD and clean the LVM physical volume.

If you want to avoid this behavior (disable automatic creation of OSD on available devices), use the `unmanaged` parameter:

```
1. ceph orch apply osd --all-available-devices --unmanaged=true
```

## Remove an OSD

```
1. ceph orch osd rm <osd_id(s)> [--replace] [--force]
```

Evacuates PGs from an OSD and removes it from the cluster.

Example:

```
1. # ceph orch osd rm 0
2. Scheduled OSD(s) for removal
```

OSDs that are not safe-to-destroy will be rejected.

You can query the state of the operation with:

```

1. # ceph orch osd rm status
2. OSD_ID HOST STATE PG_COUNT REPLACE FORCE STARTED_AT
3. 2 cephadm-dev done, waiting for purge 0 True False 2020-07-17 13:01:43.147684
4. 3 cephadm-dev draining 17 False True 2020-07-17 13:01:45.162158
5. 4 cephadm-dev started 42 False True 2020-07-17 13:01:45.162158

```

When no PGs are left on the OSD, it will be decommissioned and removed from the cluster.

#### Note

After removing an OSD, if you wipe the LVM physical volume in the device used by the removed OSD, a new OSD will be created. Read information about the `unmanaged` parameter in [Create OSDs](#).

## Stopping OSD Removal

You can stop the queued OSD removal operation with

```
1. ceph orch osd rm stop <svc_id(s)>
```

Example:

```

1. # ceph orch osd rm stop 4
2. Stopped OSD(s) removal

```

This will reset the initial state of the OSD and take it off the removal queue.

## Replace an OSD

```
1. orch osd rm <svc_id(s)> --replace [--force]
```

Example:

```

1. # ceph orch osd rm 4 --replace
2. Scheduled OSD(s) for replacement

```

This follows the same procedure as the “Remove OSD” part with the exception that the OSD is not permanently removed from the CRUSH hierarchy, but is assigned a ‘destroyed’ flag.

### PRESERVING THE OSD ID

The previously-set ‘destroyed’ flag is used to determine OSD ids that will be reused in the next OSD deployment.

If you use OSDSpecs for OSD deployment, your newly added disks will be assigned the OSD ids of their replaced counterparts, assuming the new disks still match the OSDSpecs.

For assistance in this process you can use the ‘-dry-run’ feature.

**Tip:** The name of your OSDSpec can be retrieved from `ceph orch ls`

Alternatively, you can use your OSDSpec file:

```
1. ceph orch apply osd -i <osd_spec_file> --dry-run
2. NAME          HOST   DATA    DB WAL
3. <name_of_osd_spec>  node1 /dev/vdb - -
```

If this matches your anticipated behavior, just omit the -dry-run flag to execute the deployment.

## Monitor and manager management

---

Creates or removes MONs or MGRs from the cluster. Orchestrator may return an error if it doesn’t know how to do this transition.

Update the number of monitor hosts:

```
1. ceph orch apply mon --placement=<placement> [--dry-run]
```

Where `placement` is a [Placement Specification](#).

Each host can optionally specify a network for the monitor to listen on.

Update the number of manager hosts:

```
1. ceph orch apply mgr --placement=<placement> [--dry-run]
```

Where `placement` is a [Placement Specification](#).

## Service Status

---

Print a list of services known to the orchestrator. The list can be limited to services on a particular host with the optional `-host` parameter and/or services of a particular type via optional `-type` parameter (`mon`, `osd`, `mgr`, `mds`, `rgw`):

```
1. ceph orch ls [--service_type type] [--service_name name] [--export] [--format f] [--refresh]
```

Discover the status of a particular service or daemons:

```
1. ceph orch ls --service_type type --service_name <name> [--refresh]
```

Export the service specs known to the orchestrator as yaml in format that is compatible to `ceph orch apply -i` :

```
1. ceph orch ls --export
```

## Daemon Status

Print a list of all daemons known to the orchestrator:

```
1. ceph orch ps [--hostname host] [--daemon_type type] [--service_name name] [--daemon_id id] [--format f] [--refresh]
```

Query the status of a particular service instance (mon, osd, mds, rgw). For OSDs the id is the numeric OSD ID, for MDS services it is the file system name:

```
1. ceph orch ps --daemon_type osd --daemon_id 0
```

## Deploying CephFS

In order to set up a [CephFS](#), execute:

```
1. ceph fs volume create <fs_name> <placement spec>
```

where `name` is the name of the CephFS and `placement` is a [Placement Specification](#).

This command will create the required Ceph pools, create the new CephFS, and deploy mds servers.

## Stateless services (MDS/RGW/NFS/rbd-mirror/iSCSI)

(Please note: The orchestrator will not configure the services. Please look into the corresponding documentation for service configuration details.)

The `name` parameter is an identifier of the group of instances:

- a CephFS file system for a group of MDS daemons,
- a zone name for a group of RGWs

Creating/growing/shrinking/removing services:

```

1. ceph orch apply mds <fs_name> [--placement=<placement>] [--dry-run]
   ceph orch apply rgw <realm> <zone> [--subcluster=<subcluster>] [--port=<port>] [--ssl] [--placement=
2. <placement>] [--dry-run]
3. ceph orch apply nfs <name> <pool> [--namespace=<namespace>] [--placement=<placement>] [--dry-run]
4. ceph orch rm <service_name> [--force]

```

where `placement` is a [Placement Specification](#).

e.g., `ceph orch apply mds myfs --placement="3 host1 host2 host3"`

Service Commands:

```
1. ceph orch <start|stop|restart|redeploy|reconfig> <service_name>
```

## Deploying custom containers

The orchestrator enables custom containers to be deployed using a YAML file. A corresponding [Service Specification](#) must look like:

```

1. service_type: container
2. service_id: foo
3. placement:
4. ...
5. image: docker.io/library/foo:latest
6. entrypoint: /usr/bin/foo
7. uid: 1000
8. gid: 1000
9. args:
10.    - "--net=host"
11.    - "--cpus=2"
12. ports:
13.    - 8080
14.    - 8443
15. envs:
16.    - SECRET=mypassword
17.    - PORT=8080
18.    - PUID=1000
19.    - PGID=1000
20. volume_mounts:
21.    CONFIG_DIR: /etc/foo
22. bind_mounts:
23.    - ['type=bind', 'source=lib/modules', 'destination=/lib/modules', 'ro=true']
24. dirs:
25.    - CONFIG_DIR
26. files:
27.    CONFIG_DIR/foo.conf:
28.    - refresh=true
29.    - username=xyz
30.    - "port: 1234"

```

where the properties of a service specification are:

- `service_id`

A unique name of the service.
- `image`

The name of the Docker image.
- `uid`

The UID to use when creating directories and files in the host system.
- `gid`

The GID to use when creating directories and files in the host system.
- `entrypoint`

Overwrite the default ENTRYPOINT of the image.
- `args`

A list of additional Podman/Docker command line arguments.
- `ports`

A list of TCP ports to open in the host firewall.
- `envs`

A list of environment variables.
- `bind_mounts`

When you use a bind mount, a file or directory on the host machine is mounted into the container. Relative source=... paths will be located below `/var/lib/ceph/<cluster-fsid>/<daemon-name>`.
- `volume_mounts`

When you use a volume mount, a new directory is created within Docker's storage directory on the host machine, and Docker manages that directory's contents. Relative source paths will be located below `/var/lib/ceph/<cluster-fsid>/<daemon-name>`.
- `dirs`

A list of directories that are created below `/var/lib/ceph/<cluster-fsid>/<daemon-name>`.
- `files`

A dictionary, where the key is the relative path of the file and the value the file content. The content must be double quoted when using a string. Use '\n' for line breaks in that case. Otherwise define multi-line content as list of strings. The given files will be created below the directory /var/lib/ceph/<cluster-fsid>/<daemon-name>. The absolute path of the directory where the file will be created must exist. Use the dirs property to create them if necessary.

## Service Specification

A *Service Specification* is a data structure represented as YAML to specify the deployment of services. For example:

```

1. service_type: rgw
2. service_id: realm.zone
3. placement:
4.   hosts:
5.     - host1
6.     - host2
7.     - host3
8. unmanaged: false
9. ...

```

where the properties of a service specification are:

- `service_type`

The type of the service. Needs to be either a Ceph service (`mon`, `crash`, `mds`, `mgr`, `osd` or `rbd-mirror`), a gateway (`nfs` or `rgw`), part of the monitoring stack (`alertmanager`, `grafana`, `node-exporter` or `prometheus`) or (`container`) for custom containers.

- `service_id`

The name of the service.

- `placement`

See [Placement Specification](#).

- `unmanaged`

If set to `true`, the orchestrator will not deploy nor remove any daemon associated with this service. Placement and all other properties will be ignored. This is useful, if this service should not be managed temporarily.

Each service type can have additional service specific properties.

Service specifications of type `mon`, `mgr`, and the monitoring types do not require a `service_id`.

A service of type `nfs` requires a pool name and may contain an optional namespace:

```

1. service_type: nfs
2. service_id: mynfs
3. placement:
4.   hosts:
5.     - host1
6.     - host2
7. spec:
8.   pool: mypool
9.   namespace: mynamespace

```

where `pool` is a RADOS pool where NFS client recovery data is stored and `namespace` is a RADOS namespace where NFS client recovery data is stored in the pool.

A service of type `osd` is described in [OSD Service Specification](#)

Many service specifications can be applied at once using `ceph orch apply -i` by submitting a multi-document YAML file:

```

1. cat <<EOF | ceph orch apply -i -
2. service_type: mon
3. placement:
4.   host_pattern: "mon*"
5. ---
6. service_type: mgr
7. placement:
8.   host_pattern: "mgr*"
9. ---
10. service_type: osd
11. service_id: default_drive_group
12. placement:
13.   host_pattern: "osd*"
14. data_devices:
15.   all: true
16. EOF

```

## Placement Specification

For the orchestrator to deploy a *service*, it needs to know where to deploy *daemons*, and how many to deploy. This is the role of a placement specification. Placement specifications can either be passed as command line arguments or in a YAML files.

### Explicit placements

Daemons can be explicitly placed on hosts by simply specifying them:

```
1. orch apply prometheus --placement="host1 host2 host3"
```

Or in YAML:

```

1. service_type: prometheus
2. placement:
3.   hosts:
4.     - host1
5.     - host2
6.     - host3

```

MONs and other services may require some enhanced network specifications:

```
1. orch daemon add mon --placement="myhost:[v2:1.2.3.4:3300,v1:1.2.3.4:6789]=name"
```

where `[v2:1.2.3.4:3300,v1:1.2.3.4:6789]` is the network address of the monitor and `=name` specifies the name of the new monitor.

## Placement by labels

Daemons can be explicitly placed on hosts that match a specific label:

```
1. orch apply prometheus --placement="label:mylabel"
```

Or in YAML:

```

1. service_type: prometheus
2. placement:
3.   label: "mylabel"

```

## Placement by pattern matching

Daemons can be placed on hosts as well:

```
1. orch apply prometheus --placement='myhost[1-3]'
```

Or in YAML:

```

1. service_type: prometheus
2. placement:
3.   host_pattern: "myhost[1-3]"

```

To place a service on all hosts, use `***`:

```
1. orch apply crash --placement='*''
```

Or in YAML:

```

1. service_type: node-exporter
2. placement:
3. host_pattern: "*"

```

## Setting a limit

By specifying `count`, only that number of daemons will be created:

```
1. orch apply prometheus --placement=3
```

To deploy *daemons* on a subset of hosts, also specify the count:

```
1. orch apply prometheus --placement="2 host1 host2 host3"
```

If the count is bigger than the amount of hosts, cephadm deploys one per host:

```
1. orch apply prometheus --placement="3 host1 host2"
```

results in two Prometheus daemons.

Or in YAML:

```

1. service_type: prometheus
2. placement:
3. count: 3

```

Or with hosts:

```

1. service_type: prometheus
2. placement:
3. count: 2
4. hosts:
5. - host1
6. - host2
7. - host3

```

## Updating Service Specifications

The Ceph Orchestrator maintains a declarative state of each service in a `ServiceSpec`. For certain operations, like updating the RGW HTTP port, we need to update the existing specification.

1. List the current `ServiceSpec`:

```
1. ceph orch ls --service_name=<service-name> --export > myservice.yaml
```

## 2. Update the yaml file:

```
1. vi myservice.yaml
```

## 3. Apply the new `ServiceSpec` :

```
1. ceph orch apply -i myservice.yaml [--dry-run]
```

# Configuring the Orchestrator CLI

To enable the orchestrator, select the orchestrator module to use with the `set backend` command:

```
1. ceph orch set backend <module>
```

For example, to enable the Rook orchestrator module and use it with the CLI:

```
1. ceph mgr module enable rook
2. ceph orch set backend rook
```

Check the backend is properly configured:

```
1. ceph orch status
```

# Disable the Orchestrator

To disable the orchestrator, use the empty string `""` :

```
1. ceph orch set backend ""
2. ceph mgr module disable rook
```

# Current Implementation Status

This is an overview of the current implementation status of the orchestrators.

Command	Rook	Cephadm
apply iscsi	○	✓
apply mds	✓	✓
apply mgr	○	✓

apply mon	✓	✓
apply nfs	✓	✓
apply osd	✓	✓
apply rbd-mirror	✓	✓
apply rgw	○	✓
apply container	○	✓
host add	○	✓
host ls	✓	✓
host rm	○	✓
daemon status	○	✓
daemon {stop,start,...}	○	✓
device {ident,fault}-(on,off)	○	✓
device ls	✓	✓
iscsi add	○	✓
mds add	○	✓
nfs add	○	✓
rbd-mirror add	○	✓
rgw add	○	✓
ps	✓	✓

where

- ○ = not yet implemented
- = not applicable
- ✓ = implemented

# Rook orchestrator integration

Rook (<https://rook.io/>) is an orchestration tool that can run Ceph inside a Kubernetes cluster.

The `rook` module provides integration between Ceph's orchestrator framework (used by modules such as `dashboard` to control cluster services) and Rook.

Orchestrator modules only provide services to other modules, which in turn provide user interfaces. To try out the `rook` module, you might like to use the [Orchestrator CLI](#) module.

## Requirements

- Running `ceph-mon` and `ceph-mgr` services that were set up with Rook in Kubernetes.
- Rook 0.9 or newer.

## Configuration

Because a Rook cluster's `ceph-mgr` daemon is running as a Kubernetes pod, the `rook` module can connect to the Kubernetes API without any explicit configuration.

## Development

If you are a developer, please see [Hacking on Ceph in Kubernetes with Rook](#) for instructions on setting up a development environment to work with this.

# MDS Autoscaler Module

The MDS Autoscaler Module monitors `fsmap` update notifications from the mgr daemon and takes action to spawn or kill MDS daemons for a file-system as per changes to the:

- `max_mds` config value
- `standby_count_wanted` config value
- standby promotions to active MDS state in case of active MDS rank death

Bumping up the `max_mds` config option value causes a standby mds to be promoted to hold an active rank. This leads to a drop in standby mds count. The MDS Autoscaler module detects this deficit and the orchestrator module is notified about the required MDS count. The orchestrator back-end then takes necessary measures to spawn standby MDSS.

Dropping the `max_mds` config option causes the orchestrator back-end to kill standby mds to achieve the new reduced count. Preferably standby mds are chosen to be killed when the `max_mds` count is dropped.

An increment and decrement of the `standby_count_wanted` config option value has a similar effect on the total MDS count. The orchestrator is notified about the change and necessary action to spawn or kill standby MDSS is taken.

A death of an active MDS rank also causes promotion of a standby mds to occupy the required active rank. The MDS Autoscaler notices the change in the standby mds count and a message is passed to the orchestrator to maintain the necessary MDS count.

NOTE: There is no CLI associated with the MDS Autoscaler Module.

# Ceph Dashboard

## Overview

The Ceph Dashboard is a built-in web-based Ceph management and monitoring application through which you can inspect and administer various aspects and resources within the cluster. It is implemented as a [Ceph Manager Daemon](#) module.

The original Ceph Dashboard that was shipped with Ceph Luminous started out as a simple read-only view into run-time information and performance data of Ceph clusters. It used a very simple architecture to achieve the original goal. However, there was growing demand for richer web-based management capabilities, to make it easier to administer Ceph for users that prefer a WebUI over the CLI.

The new [Ceph Dashboard](#) module adds web-based monitoring and administration to the Ceph Manager. The architecture and functionality of this new module are derived from and inspired by the [openATTIC Ceph management and monitoring tool](#). Development is actively driven by the openATTIC team at [SUSE](#), with support from companies including [Red Hat](#) and members of the Ceph community.

The dashboard module's backend code uses the CherryPy framework and implements a custom REST API. The WebUI implementation is based on Angular/TypeScript and includes both functionality from the original dashboard and new features originally developed for the standalone version of openATTIC. The Ceph Dashboard module is implemented as an application that provides a graphical representation of information and statistics through a web server hosted by [ceph-mgr](#).

## Feature Overview

The dashboard provides the following features:

- **Multi-User and Role Management:** The dashboard supports multiple user accounts with different permissions (roles). User accounts and roles can be managed via both the command line and the WebUI. The dashboard supports various methods to enhance password security. Password complexity rules may be configured, requiring users to change their password after the first login or after a configurable time period. See [User and Role Management](#) for details.
- **Single Sign-On (SSO):** The dashboard supports authentication via an external identity provider using the SAML 2.0 protocol. See [Enabling Single Sign-On \(SSO\)](#) for details.
- **SSL/TLS support:** All HTTP communication between the web browser and the dashboard is secured via SSL. A self-signed certificate can be created with a built-in command, but it's also possible to import custom certificates signed and issued

by a CA. See [SSL/TLS Support](#) for details.

- **Auditing:** The dashboard backend can be configured to log all `PUT`, `POST` and `DELETE` API requests in the Ceph audit log. See [Auditing API Requests](#) for instructions on how to enable this feature.
- **Internationalization (I18N):** The language used for dashboard text can be selected at run-time.

The Ceph Dashboard offers the following monitoring and management capabilities:

- **Overall cluster health:** Display performance and capacity metrics as well as cluster status.
- **Embedded Grafana Dashboards:** Ceph Dashboard [Grafana](#) dashboards may be embedded in external applications and web pages to surface information and performance metrics gathered by the [Prometheus Module](#) module. See [Enabling the Embedding of Grafana Dashboards](#) for details on how to configure this functionality.
- **Cluster logs:** Display the latest updates to the cluster's event and audit log files. Log entries can be filtered by priority, date or keyword.
- **Hosts:** Display a list of all cluster hosts along with their storage drives, which services are running, and which version of Ceph is installed.
- **Performance counters:** Display detailed service-specific statistics for each running service.
- **Monitors:** List all Mons, their quorum status, and open sessions.
- **Monitoring:** Enable creation, re-creation, editing, and expiration of Prometheus' silences, list the alerting configuration and all configured and firing alerts. Show notifications for firing alerts.
- **Configuration Editor:** Display all available configuration options, their descriptions, types, default and currently set values. These may be edited as well.
- **Pools:** List Ceph pools and their details (e.g. applications, pg-autoscaling, placement groups, replication size, EC profile, CRUSH rulesets, quotas etc.)
- **OSDs:** List OSDs, their status and usage statistics as well as detailed information like attributes (OSD map), metadata, performance counters and usage histograms for read/write operations. Mark OSDs up/down/out, purge and reweight OSDs, perform scrub operations, modify various scrub-related configuration options, select profiles to adjust the level of backfilling activity. List all drives associated with an OSD. Set and change the device class of an OSD, display and sort OSDs by device class. Deploy OSDs on new drives and hosts.
- **Device management:** List all hosts known by the orchestrator. List all drives

attached to a host and their properties. Display drive health predictions and SMART data. Blink enclosure LEDs.

- **iSCSI:** List all hosts that run the TCMU runner service, display all images and their performance characteristics (read/write ops, traffic). Create, modify, and delete iSCSI targets (via `ceph-iscsi`). Display the iSCSI gateway status and info about active initiators. See [Enabling iSCSI Management](#) for instructions on how to configure this feature.
- **RBD:** List all RBD images and their properties (size, objects, features). Create, copy, modify and delete RBD images (incl. snapshots) and manage RBD namespaces. Define various I/O or bandwidth limitation settings on a global, per-pool or per-image level. Create, delete and rollback snapshots of selected images, protect/unprotect these snapshots against modification. Copy or clone snapshots, flatten cloned images.
- **RBD mirroring:** Enable and configure RBD mirroring to a remote Ceph server. List active daemons and their status, pools and RBD images including sync progress.
- **CephFS:** List active file system clients and associated pools, including usage statistics. Evict active CephFS clients. Manage CephFS quotas and snapshots. Browse a CephFS directory structure.
- **Object Gateway:** List all active object gateways and their performance counters. Display and manage (add/edit/delete) object gateway users and their details (e.g. quotas) as well as the users' buckets and their details (e.g. placement targets, owner, quotas, versioning, multi-factor authentication). See [Enabling the Object Gateway Management Frontend](#) for configuration instructions.
- **NFS:** Manage NFS exports of CephFS file systems and RGW S3 buckets via NFS Ganesha. See [NFS-Ganesha Management](#) for details on how to enable this functionality.
- **Ceph Manager Modules:** Enable and disable Ceph Manager modules, manage module-specific configuration settings.

## Overview of the Dashboard Landing Page

Displays overall cluster status, performance, and capacity metrics. Shows instant feedback for changes in the cluster and provides easy access to subpages of the dashboard.

## Status

- **Cluster Status:** Displays overall cluster health. In case of any error it displays a short description of the error and provides a link to the logs.
- **Hosts:** Displays the total number of hosts associated to the cluster and links to a subpage that lists and describes each.

- **Monitors:** Displays mons and their quorum status and open sessions. Links to a subpage that lists and describes each.
- **OSDs:** Displays object storage daemons (ceph-osds) and the numbers of OSDs running (up), in service (in), and out of the cluster (out). Provides links to subpages providing a list of all OSDs and related management actions.
- **Managers:** Displays active and standby Ceph Manager daemons (ceph-mgr).
- **Object Gateway:** Displays active object gateways (RGWs) and provides links to subpages that list all object gateway daemons.
- **Metadata Servers:** Displays active and standby CephFS metadata service daemons (ceph-mds).
- **iSCSI Gateways:** Displays iSCSI gateways available, active (up), and inactive (down). Provides a link to a subpage showing a list of all iSCSI Gateways.

## Capacity

- **Raw Capacity:** Displays the capacity used out of the total physical capacity provided by storage nodes (OSDs).
- **Objects:** Displays the number and status of RADOS objects including the percentages of healthy, misplaced, degraded, and unfound objects.
- **PG Status:** Displays the total number of placement groups and their status, including the percentage clean, working, warning, and unknown.
- **Pools:** Displays pools and links to a subpage listing details.
- **PGs per OSD:** Displays the number of placement groups assigned to object storage daemons.

## Performance

- **Client READ/Write:** Displays an overview of client input and output operations.
- **Client Throughput:** Displays the data transfer rates to and from Ceph clients.
- **Recovery throughput:** Displays rate of cluster healing and balancing operations.
- **Scrubbing:** Displays light and deep scrub status.

## Supported Browsers

Ceph Dashboard is primarily tested and developed using the following web browsers:

Browser	Versions

Chrome and Chromium based browsers	latest 2 major versions
Firefox	latest 2 major versions
Firefox ESR	latest major version

While Ceph Dashboard might work in older browsers, we cannot guarantee compatibility and recommend keeping your browser up to date.

## Enabling

If you have installed `ceph-mgr-dashboard` from distribution packages, the package management system should take care of installing all required dependencies.

If you're building Ceph from source and want to start the dashboard from your development environment, please see the files `README.rst` and `HACKING.rst` in the source directory `src/pybind/mgr/dashboard`.

Within a running Ceph cluster, the Ceph Dashboard is enabled with:

```
1. $ ceph mgr module enable dashboard
```

## Configuration

### SSL/TLS Support

All HTTP connections to the dashboard are secured with SSL/TLS by default.

To get the dashboard up and running quickly, you can generate and install a self-signed certificate:

```
1. $ ceph dashboard create-self-signed-cert
```

Note that most web browsers will complain about self-signed certificates and require explicit confirmation before establishing a secure connection to the dashboard.

To properly secure a deployment and to remove the warning, a certificate that is issued by a certificate authority (CA) should be used.

For example, a key pair can be generated with a command similar to:

```
1. $ openssl req -new -nodes -x509 \
2.   -subj "/O=IT/CN=ceph-mgr-dashboard" -days 3650 \
3.   -keyout dashboard.key -out dashboard.crt -extensions v3_ca
```

The `dashboard.crt` file should then be signed by a CA. Once that is done, you can enable it for Ceph manager instances by running the following commands:

```
1. $ ceph dashboard set-ssl-certificate -i dashboard.crt  
2. $ ceph dashboard set-ssl-certificate-key -i dashboard.key
```

If unique certificates are desired for each manager instance, the name of the instance can be included as follows (where `$name` is the name of the `ceph-mgr` instance, usually the hostname):

```
1. $ ceph dashboard set-ssl-certificate $name -i dashboard.crt  
2. $ ceph dashboard set-ssl-certificate-key $name -i dashboard.key
```

SSL can also be disabled by setting this configuration value:

```
1. $ ceph config set mgr mgr/dashboard/ssl false
```

This might be useful if the dashboard will be running behind a proxy which does not support SSL for its upstream servers or other situations where SSL is not wanted or required. See [Proxy Configuration](#) for more details.

#### Warning

Use caution when disabling SSL as usernames and passwords will be sent to the dashboard unencrypted.

#### Note

You must restart Ceph manager processes after changing the SSL certificate and key. This can be accomplished by either running `ceph mgr fail mgr` or by disabling and re-enabling the dashboard module (which also triggers the manager to respawn itself):

```
1. $ ceph mgr module disable dashboard  
2. $ ceph mgr module enable dashboard
```

## Host Name and Port

Like most web applications, the dashboard binds to a TCP/IP address and TCP port.

By default, the `ceph-mgr` daemon hosting the dashboard (i.e., the currently active manager) will bind to TCP port 8443 or 8080 when SSL is disabled.

If no specific address has been configured, the web app will bind to `::`, which corresponds to all available IPv4 and IPv6 addresses.

These defaults can be changed via the configuration key facility on a cluster-wide level (so they apply to all manager instances) as follows:

```

1. $ ceph config set mgr mgr/dashboard/server_addr $IP
2. $ ceph config set mgr mgr/dashboard/server_port $PORT
3. $ ceph config set mgr mgr/dashboard/ssl_server_port $PORT

```

Since each `ceph-mgr` hosts its own instance of the dashboard, it may be necessary to configure them separately. The IP address and port for a specific manager instance can be changed with the following commands:

```

1. $ ceph config set mgr mgr/dashboard/$name/server_addr $IP
2. $ ceph config set mgr mgr/dashboard/$name/server_port $PORT
3. $ ceph config set mgr mgr/dashboard/$name/ssl_server_port $PORT

```

Replace `$name` with the ID of the ceph-mgr instance hosting the dashboard.

#### Note

The command `ceph mgr services` will show you all endpoints that are currently configured. Look for the `dashboard` key to obtain the URL for accessing the dashboard.

## Username and Password

In order to be able to log in, you need to create a user account and associate it with at least one role. We provide a set of predefined *system roles* that you can use. For more details please refer to the [User and Role Management](#) section.

To create a user with the administrator role you can use the following commands:

```
1. $ ceph dashboard ac-user-create <username> <password> administrator
```

## Accessing the Dashboard

You can now access the dashboard using your (JavaScript-enabled) web browser, by pointing it to any of the host names or IP addresses and the selected TCP port where a manager instance is running: e.g., `http(s)://<$IP>:<$PORT>/`.

The dashboard page displays and requests a previously defined username and password.

## Enabling the Object Gateway Management Frontend

To use the Object Gateway management functionality of the dashboard, you will need to provide the login credentials of a user with the `system` flag enabled. If you do not have a `system` user already, you must create one:

```

1. $ radosgw-admin user create --uid=<user_id> --display-name=<display_name> \
2.   --system

```

Take note of the keys `access_key` and `secret_key` in the output.

To obtain the credentials of an existing user via `radosgw-admin`:

```
1. $ radosgw-admin user info --uid=<user_id>
```

Finally, provide the credentials to the dashboard:

```
1. $ ceph dashboard set-rgw-api-access-key <access_key>
2. $ ceph dashboard set-rgw-api-secret-key <secret_key>
```

In a simple configuration with a single RGW endpoint, this is all you have to do to get the Object Gateway management functionality working. The dashboard will try to automatically determine the host and port from the Ceph Manager's service map.

If multiple zones are used, it will automatically determine the host within the master zone group and master zone. This should be sufficient for most setups, but in some circumstances you might want to set the host and port manually:

```
1. $ ceph dashboard set-rgw-api-host <host>
2. $ ceph dashboard set-rgw-api-port <port>
```

In addition to the settings mentioned so far, the following settings do also exist and you may find yourself in the situation that you have to use them:

```
1. $ ceph dashboard set-rgw-api-scheme <scheme> # http or https
2. $ ceph dashboard set-rgw-api-admin-resource <admin_resource>
3. $ ceph dashboard set-rgw-api-user-id <user_id>
```

If you are using a self-signed certificate in your Object Gateway setup, you should disable certificate verification in the dashboard to avoid refused connections, e.g. caused by certificates signed by unknown CA or not matching the host name:

```
1. $ ceph dashboard set-rgw-api-ssl-verify False
```

If the Object Gateway takes too long to process requests and the dashboard runs into timeouts, you can set the timeout value to your needs:

```
1. $ ceph dashboard set-rest-requests-timeout <seconds>
```

The default value is 45 seconds.

## Enabling iSCSI Management

The Ceph Dashboard can manage iSCSI targets using the REST API provided by the `rbd-target-api` service of the [Ceph iSCSI Gateway](#). Please make sure that it is installed and

enabled on the iSCSI gateways.

#### Note

The iSCSI management functionality of Ceph Dashboard depends on the latest version 3 of the [ceph-iscsi](#) project. Make sure that your operating system provides the correct version, otherwise the dashboard will not enable the management features.

If the `ceph-iscsi` REST API is configured in HTTPS mode and its using a self-signed certificate, you need to configure the dashboard to avoid SSL certificate verification when accessing ceph-iscsi API.

To disable API SSL verification run the following command:

```
1. $ ceph dashboard set-iscsi-api-ssl-verification false
```

The available iSCSI gateways must be defined using the following commands:

```
1. $ ceph dashboard iscsi-gateway-list
2. $ ceph dashboard iscsi-gateway-add <scheme>://<username>:<password>@<host>[:port]
3. $ ceph dashboard iscsi-gateway-rm <gateway_name>
```

## Enabling the Embedding of Grafana Dashboards

Grafana pulls data from [Prometheus](#). Although Grafana can use other data sources, the Grafana dashboards we provide contain queries that are specific to Prometheus. Our Grafana dashboards therefore require Prometheus as the data source. The Ceph [Prometheus Module](#) module exports its data in the Prometheus exposition format. These Grafana dashboards rely on metric names from the Prometheus module and [Node exporter](#). The Node exporter is a separate application that provides machine metrics.

#### Note

Prometheus' security model presumes that untrusted users have access to the Prometheus HTTP endpoint and logs. Untrusted users have access to all the (meta)data Prometheus collects that is contained in the database, plus a variety of operational and debugging information.

However, Prometheus' HTTP API is limited to read-only operations. Configurations can *not* be changed using the API and secrets are not exposed. Moreover, Prometheus has some built-in measures to mitigate the impact of denial of service attacks.

Please see Prometheus' Security model

<https://prometheus.io/docs/operating/security/>; for more detailed information.

## Installation and Configuration using cephadm

Grafana and Prometheus can be installed using [Cephadm](#). They will automatically be

configured by `cephadm`. Please see [Monitoring Stack with Cephadm](#) documentation for more details on how to use `cephadm` for installing and configuring Prometheus and Grafana.

## Manual Installation and Configuration

The following process describes how to configure Grafana and Prometheus manually. After you have installed Prometheus, Grafana, and the Node exporter on appropriate hosts, proceed with the following steps.

1. Enable the Ceph Exporter which comes as Ceph Manager module by running:

```
1. $ ceph mgr module enable prometheus
```

More details can be found in the documentation of the [Prometheus Module](#).

2. Add the corresponding scrape configuration to Prometheus. This may look like:

```
1. global:
2.   scrape_interval: 5s
3.
4. scrape_configs:
5.   - job_name: 'prometheus'
6.     static_configs:
7.       - targets: ['localhost:9090']
8.   - job_name: 'ceph'
9.     static_configs:
10.      - targets: ['localhost:9283']
11.   - job_name: 'node-exporter'
12.     static_configs:
13.       - targets: ['localhost:9100']
```

### Note

Please note that in the above example, Prometheus is configured to scrape data from itself (port 9090), the Ceph manager module prometheus (port 9283), which exports Ceph internal data, and the Node Exporter (port 9100), which provides OS and hardware metrics for each host.

Depending on your configuration, you may need to change the hostname in or add additional configuration entries for the Node Exporter. It is unlikely that you will need to change the default TCP ports.

Moreover, you don't *need* to have more than one target for Ceph specific data, provided by the prometheus mgr module. But it is recommended to configure Prometheus to scrape Ceph specific data from all existing Ceph managers. This enables a built-in high availability mechanism, so that services run on a manager host will be restarted automatically on a different manager host if one Ceph Manager goes down.

3. Add Prometheus as data source to Grafana [using the Grafana Web UI](#).
4. Install the vonage-status-panel and grafana-piechart-panel plugins using:

```
1. grafana-cli plugins install vonage-status-panel
2. grafana-cli plugins install grafana-piechart-panel
```

5. Add Dashboards to Grafana:

Dashboards can be added to Grafana by importing dashboard JSON files. Use the following command to download the JSON files:

```
wget https://raw.githubusercontent.com/ceph/ceph/master/monitoring/grafana/dashboards/<Dashboard-name>.json
```

You can find various dashboard JSON files [here](#).

For Example, for ceph-cluster overview you can use:

```
1. wget https://raw.githubusercontent.com/ceph/ceph/master/monitoring/grafana/dashboards/ceph-cluster.json
```

You may also author your own dashboards.

6. Configure anonymous mode in `/etc/grafana/grafana.ini` :

```
1. [auth.anonymous]
2. enabled = true
3. org_name = Main Org.
4. org_role = Viewer
```

In newer versions of Grafana (starting with 6.2.0-beta1) a new setting named `allow_embedding` has been introduced. This setting must be explicitly set to `true` for the Grafana integration in Ceph Dashboard to work, as the default is `false`.

```
1. [security]
2. allow_embedding = true
```

## Enabling RBD-Image monitoring

Monitoring of RBD images is disabled by default, as it can significantly impact performance. For more information please see [RBD IO statistics](#). When disabled, the overview and details dashboards will be empty in Grafana and metrics will not be visible in Prometheus.

## Configuring Dashboard

After you have set up Grafana and Prometheus, you will need to configure the

connection information that the Ceph Dashboard will use to access Grafana.

Tell the dashboard the URL for the deployed Grafana instance:

```
1. $ ceph dashboard set-grafana-api-url <grafana-server-url> # default: ''
```

The format of url is : <protocol>:<IP-address>:<port>

#### Note

The Ceph Dashboard embeds Grafana dashboards via `iframe` HTML elements. If Grafana is configured without SSL/TLS support, most browsers will block the embedding of insecure content if SSL support is enabled for the dashboard (which is the default). If you can't see the embedded Grafana dashboards after enabling them as outlined above, check your browser's documentation on how to unblock mixed content. Alternatively, consider enabling SSL/TLS support in Grafana.

If you are using a self-signed certificate for Grafana, disable certificate verification in the dashboard to avoid refused connections, which can be a result of certificates signed by an unknown CA or that do not matchn the host name:

```
1. $ ceph dashboard set-grafana-api-ssl-verify False
```

You can also access Grafana directly to monitor your cluster.

#### Note

Ceph Dashboard configuration information can also be unset. For example, to clear the Grafana API URL we configured above:

```
1. $ ceph dashboard reset-grafana-api-url
```

## Enabling Single Sign-On (SSO)

The Ceph Dashboard supports external authentication of users via the [SAML 2.0](#) protocol. You need to first create user accounts and associate them with desired roles, as authorization is performed by the Dashboard. However, the authentication process can be performed by an existing Identity Provider (IdP).

#### Note

Ceph Dashboard SSO support relies on onelogin's [python-saml](#) library. Please ensure that this library is installed on your system, either by using your distribution's package management or via Python's pip installer.

To configure SSO on Ceph Dashboard, you should use the following command:

```
$ ceph dashboard sso setup saml2 <ceph_dashboard_base_url> <idp_metadata> {<idp_username_attribute>}
1. {<idp_entity_id>} {<sp_x_509_cert>} {<sp_private_key>}
```

Parameters:

- **<ceph\_dashboard\_base\_url>**: Base URL where Ceph Dashboard is accessible (e.g., <https://cephdashboard.local>)
- **<idp\_metadata>**: URL to remote (http://, https://) or local (file://) path or content of the IdP metadata XML (e.g., <https://myidp/metadata>, file:///home/myuser/metadata.xml).
- **<idp\_username\_attribute> (optional)**: Attribute that should be used to get the username from the authentication response. Defaults to uid.
- **<idp\_entity\_id> (optional)**: Use this when more than one entity id exists on the IdP metadata.
- **<sp\_x\_509\_cert> / <sp\_private\_key> (optional)**: File path of the certificate that should be used by Ceph Dashboard (Service Provider) for signing and encryption.

Note

The issuer value of SAML requests will follow this pattern:

**<ceph\_dashboard\_base\_url>/auth/saml2/metadata**

To display the current SAML 2.0 configuration, use the following command:

```
1. $ ceph dashboard sso show saml2
```

Note

For more information about onelogin\_settings, please check the [onelogin documentation](#).

To disable SSO:

```
1. $ ceph dashboard sso disable
```

To check if SSO is enabled:

```
1. $ ceph dashboard sso status
```

To enable SSO:

```
1. $ ceph dashboard sso enable saml2
```

## Enabling Prometheus Alerting

To use Prometheus for alerting you must define [alerting rules](#). These are managed by the [Alertmanager](#). If you are not yet using the Alertmanager, [install it](#) as it receives and manages alerts from Prometheus.

Alertmanager capabilities can be consumed by the dashboard in three different ways:

1. Use the notification receiver of the dashboard.
2. Use the Prometheus Alertmanager API.
3. Use both sources simultaneously.

All three methods notify you about alerts. You won't be notified twice if you use both sources, but you need to consume at least the Alertmanager API in order to manage silences.

### 1. Use the notification receiver of the dashboard

This allows you to get notifications as [configured](#) from the Alertmanager. You will get notified inside the dashboard once a notification is sent out, but you are not able to manage alerts.

Add the dashboard receiver and the new route to your Alertmanager configuration. This should look like:

```

1. route:
2.   receiver: 'ceph-dashboard'
3. ...
4. receivers:
5.   - name: 'ceph-dashboard'
6.   webhook_configs:
7.     - url: '<url-to-dashboard>/api/prometheus_receiver'
```

Ensure that the Alertmanager considers your SSL certificate in terms of the dashboard as valid. For more information about the correct configuration checkout the [<http\\_config> documentation](#).

### 1. Use the API of Prometheus and the Alertmanager

This allows you to manage alerts and silences and will enable the "Active Alerts", "All Alerts" as well as the "Silences" tabs in the "Monitoring" section of the "Cluster" menu entry.

Alerts can be sorted by name, job, severity, state and start time. Unfortunately it's not possible to know when an alert was sent out through a notification by the Alertmanager based on your configuration, that's why the dashboard will notify the user on any visible change to an alert and will notify the changed alert.

Silences can be sorted by id, creator, status, start, updated and end time. Silences can be created in various ways, it's also possible to expire them.

1. Create from scratch
2. Based on a selected alert
3. Recreate from expired silence
4. Update a silence (which will recreate and expire it (default Alertmanager behaviour))

To use it, specify the host and port of the Alertmanager server:

```
1. $ ceph dashboard set-alertmanager-api-host <alertmanager-host:port> # default: ''
```

For example:

```
1. $ ceph dashboard set-alertmanager-api-host 'http://localhost:9093'
```

To be able to see all configured alerts, you will need to configure the URL to the Prometheus API. Using this API, the UI will also help you in verifying that a new silence will match a corresponding alert.

```
1. $ ceph dashboard set-prometheus-api-host <prometheus-host:port> # default: ''
```

For example:

```
1. $ ceph dashboard set-prometheus-api-host 'http://localhost:9090'
```

After setting up the hosts, refresh your browser's dashboard window or tab.

## 1. Use both methods

The behaviors of both methods are configured in a way that they should not disturb each other, through annoying duplicated notifications may pop up.

If you are using a self-signed certificate in your Prometheus or your Alertmanager setup, you should disable certificate verification in the dashboard to avoid refused connections caused by certificates signed by an unknown CA or that do not match the host name.

- For Prometheus:

```
1. $ ceph dashboard set-prometheus-api-ssl-verify False
```

- For Alertmanager:

```
1. $ ceph dashboard set-alertmanager-api-ssl-verify False
```

# User and Role Management

## Password Policy

By default the password policy feature is enabled, which includes the following checks:

- Is the password longer than N characters?
- Are the old and new password the same?

The password policy feature can be switched on or off completely:

```
1. $ ceph dashboard set-pwd-policy-enabled <true|false>
```

The following individual checks can also be switched on or off:

```
1. $ ceph dashboard set-pwd-policy-check-length-enabled <true|false>
2. $ ceph dashboard set-pwd-policy-check-oldpwd-enabled <true|false>
3. $ ceph dashboard set-pwd-policy-check-username-enabled <true|false>
4. $ ceph dashboard set-pwd-policy-check-exclusion-list-enabled <true|false>
5. $ ceph dashboard set-pwd-policy-check-complexity-enabled <true|false>
6. $ ceph dashboard set-pwd-policy-check-sequential-chars-enabled <true|false>
7. $ ceph dashboard set-pwd-policy-check-repetitive-chars-enabled <true|false>
```

Additionally the following options are available to configure password policy.

- Minimum password length (defaults to 8):

```
1. $ ceph dashboard set-pwd-policy-min-length <N>
```

- Minimum password complexity (defaults to 10):

```
1. $ ceph dashboard set-pwd-policy-min-complexity <N>
```

Password complexity is calculated by classifying each character in the password. The complexity count starts by 0. A character is rated by the following rules in the given order.

- Increase by 1 if the character is a digit.
- Increase by 1 if the character is a lower case ASCII character.
- Increase by 2 if the character is an upper case ASCII character.
- Increase by 3 if the character is a special character like !#\$%&'()\*+,-./:;<=>?  
@[\]^\_`{|}~ .
- Increase by 5 if the character has not been classified by one of the previous rules.
- A list of comma separated words that are not allowed to be used in a password:

```
1. $ ceph dashboard set-pwd-policy-exclusion-list <word>[,...]
```

## User Accounts

The Ceph Dashboard supports multiple user accounts. Each user account consists of a username, a password (stored in encrypted form using `bcrypt`), an optional name, and an optional email address.

If a new user is created via the Web UI, it is possible to set an option that the user must assign a new password when they log in for the first time.

User accounts are stored in the monitors' configuration database, and are available to all `ceph-mgr` instances.

We provide a set of CLI commands to manage user accounts:

- *Show User(s):*

```
1. $ ceph dashboard ac-user-show [<username>]
```

- *Create User:*

```
$ ceph dashboard ac-user-create [--enabled] [--force-password] [--pwd_update_required] <username>
1. [<password>] [<rolename>] [<name>] [<email>] [<pwd_expiration_date>]
```

To bypass password policy checks use the `force-password` option. Add the option `pwd_update_required` so that a newly created user has to change their password after the first login.

- *Delete User:*

```
1. $ ceph dashboard ac-user-delete <username>
```

- *Change Password:*

```
1. $ ceph dashboard ac-user-set-password [--force-password] <username> <password>
```

- *Change Password Hash:*

```
1. $ ceph dashboard ac-user-set-password-hash <username> <hash>
```

The hash must be a bcrypt hash and salt, e.g.

`$2b$12$Pt3Vq/rDt2y9glTPSV.VFegiLkQeIpddtkhoFetNApYmIJ0Y8gau2`. This can be used to import users from an external database.

- *Modify User (name, and email):*

```
1. $ ceph dashboard ac-user-set-info <username> <name> <email>
```

- *Disable User:*

```
1. $ ceph dashboard ac-user-disable <username>
```

- *Enable User:*

```
1. $ ceph dashboard ac-user-enable <username>
```

## User Roles and Permissions

User accounts are associated with a set of roles that define which dashboard functionality can be accessed.

The Dashboard functionality/modules are grouped within a *security scope*. Security scopes are predefined and static. The current available security scopes are:

- **hosts**: includes all features related to the `Hosts` menu entry.
- **config-opt**: includes all features related to management of Ceph configuration options.
- **pool**: includes all features related to pool management.
- **osd**: includes all features related to OSD management.
- **monitor**: includes all features related to monitor management.
- **rbd-image**: includes all features related to RBD image management.
- **rbd-mirroring**: includes all features related to RBD mirroring management.
- **iscsi**: includes all features related to iSCSI management.
- **rgw**: includes all features related to RADOS Gateway (RGW) management.
- **cephfs**: includes all features related to CephFS management.
- **manager**: include all features related to Ceph Manager management.
- **log**: include all features related to Ceph logs management.
- **grafana**: include all features related to Grafana proxy.
- **prometheus**: include all features related to Prometheus alert management.
- **dashboard-settings**: allows to change dashboard settings.

A *role* specifies a set of mappings between a *security scope* and a set of *permissions*. There are four types of permissions:

- **read**
- **create**
- **update**
- **delete**

See below for an example of a role specification, in the form of a Python dictionary:

```

1. # example of a role
2. {
3.     'role': 'my_new_role',
4.     'description': 'My new role',
5.     'scopes_permissions': {
6.         'pool': ['read', 'create'],
7.         'rbd-image': ['read', 'create', 'update', 'delete']
8.     }
9. }
```

The above role dictates that a user has *read* and *create* permissions for features related to pool management, and has full permissions for features related to RBD image management.

The Dashboard provides a set of predefined roles that we call *system roles*, which can be used right away by a fresh Ceph Dashboard installation.

The list of system roles are:

- **administrator**: allows full permissions for all security scopes.
- **read-only**: allows *read* permission for all security scopes except dashboard settings.
- **block-manager**: allows full permissions for *rbd-image*, *rbd-mirroring*, and *iscsi* scopes.
- **rgw-manager**: allows full permissions for the *rgw* scope
- **cluster-manager**: allows full permissions for the *hosts*, *osd*, *monitor*, *manager*, and *config-opt* scopes.
- **pool-manager**: allows full permissions for the *pool* scope.
- **cephfs-manager**: allows full permissions for the *cephfs* scope.

The list of available roles can be retrieved with the following command:

```
1. $ ceph dashboard ac-role-show [<rolename>]
```

You can also use the CLI to create new roles. The available commands are the following:

- *Create Role*:

```
1. $ ceph dashboard ac-role-create <rolename> [<description>]
```

- *Delete Role*:

```
1. $ ceph dashboard ac-role-delete <rolename>
```

- Add Scope Permissions to Role:

```
1. $ ceph dashboard ac-role-add-scope-perms <rolename> <scopename> <permission> [<permission>...]
```

- Delete Scope Permission from Role:

```
1. $ ceph dashboard ac-role-del-scope-perms <rolename> <scopename>
```

To assign roles to users, the following commands are available:

- Set User Roles:

```
1. $ ceph dashboard ac-user-set-roles <username> <rolename> [<rolename>...]
```

- Add Roles To User:

```
1. $ ceph dashboard ac-user-add-roles <username> <rolename> [<rolename>...]
```

- Delete Roles from User:

```
1. $ ceph dashboard ac-user-del-roles <username> <rolename> [<rolename>...]
```

## Example of User and Custom Role Creation

In this section we show a complete example of the commands that create a user account that can manage RBD images, view and create Ceph pools, and has read-only access to other scopes.

1. Create the user:

```
1. $ ceph dashboard ac-user-create bob mypassword
```

2. Create role and specify scope permissions:

```
1. $ ceph dashboard ac-role-create rbd/pool-manager
2. $ ceph dashboard ac-role-add-scope-perms rbd/pool-manager rbd-image read create update delete
3. $ ceph dashboard ac-role-add-scope-perms rbd/pool-manager pool read create
```

3. Associate roles to user:

```
1. $ ceph dashboard ac-user-set-roles bob rbd/pool-manager read-only
```

# Proxy Configuration

In a Ceph cluster with multiple `ceph-mgr` instances, only the dashboard running on the currently active `ceph-mgr` daemon will serve incoming requests. Connections to the dashboard's TCP port on standby `ceph-mgr` instances will receive an HTTP redirect (303) to the active manager's dashboard URL. This enables you to point your browser to any `ceph-mgr` instance in order to access the dashboard.

If you want to establish a fixed URL to reach the dashboard or if you don't want to allow direct connections to the manager nodes, you could set up a proxy that automatically forwards incoming requests to the active `ceph-mgr` instance.

## Configuring a URL Prefix

If you are accessing the dashboard via a reverse proxy, you may wish to service it under a URL prefix. To get the dashboard to use hyperlinks that include your prefix, you can set the `url_prefix` setting:

```
1. ceph config set mgr mgr/dashboard/url_prefix $PREFIX
```

so you can access the dashboard at `http://$IP:$PORT/$PREFIX/`.

## Disable the redirection

If the dashboard is behind a load-balancing proxy like `HAProxy` you might want to disable redirection to prevent situations in which internal (unresolvable) URLs are published to the frontend client. Use the following command to get the dashboard to respond with an HTTP error (500 by default) instead of redirecting to the active dashboard:

```
1. $ ceph config set mgr mgr/dashboard/standby_behaviour "error"
```

To reset the setting to default redirection, use the following command:

```
1. $ ceph config set mgr mgr/dashboard/standby_behaviour "redirect"
```

## Configure the error status code

When redirection is disabled, you may want to customize the HTTP status code of standby dashboards. To do so you need to run the command:

```
1. $ ceph config set mgr mgr/dashboard/standby_error_status_code 503
```

## HAProxy example configuration

Below you will find an example configuration for SSL/TLS passthrough using [HAProxy](#).

Please note that this configuration works under the following conditions. If the dashboard fails over, the front-end client might receive a HTTP redirect (303) response and will be redirected to an unresolvable host. This happens when failover occurs between two HAProxy health checks. In this situation the previously active dashboard node will now respond with a 303 which points to the new active node. To prevent that situation you should consider disabling redirection on standby nodes.

```

1. defaults
2.   log global
3.   option log-health-checks
4.   timeout connect 5s
5.   timeout client 50s
6.   timeout server 450s
7.
8. frontend dashboard_front
9.   mode http
10.  bind *:80
11.  option httplog
12.  redirect scheme https code 301 if !{ ssl_fc }
13.
14. frontend dashboard_front_ssl
15.   mode tcp
16.   bind *:443
17.   option tcplog
18.   default_backend dashboard_back_ssl
19.
20. backend dashboard_back_ssl
21.   mode tcp
22.   option httpchk GET /
23.   http-check expect status 200
24.   server x <HOST>:<PORT> ssl check verify none
25.   server y <HOST>:<PORT> ssl check verify none
26.   server z <HOST>:<PORT> ssl check verify none

```

## Auditing API Requests

The REST API can log PUT, POST and DELETE requests to the Ceph audit log. This feature is disabled by default, but can be enabled with the following command:

```
1. $ ceph dashboard set-audit-api-enabled <true|false>
```

If enabled, the following parameters are logged per each request:

- from - The origin of the request, e.g. [https://\[::1\]:44410](https://[::1]:44410)
- path - The REST API path, e.g. /api/auth

- method - e.g. PUT, POST or DELETE
- user - The name of the user, otherwise 'None'

The logging of the request payload (the arguments and their values) is enabled by default. Execute the following command to disable this behaviour:

```
1. $ ceph dashboard set-audit-api-log-payload <true|false>
```

A log entry may look like this:

```
2018-10-22 15:27:01.302514 mgr.x [INF] [DASHBOARD] from='https://[::ffff:127.0.0.1]:37022'
path='/api/rgw/user/klaus' method='PUT' user='admin' params='{"max_buckets": "1000", "display_name": "Klaus
1. Mustermann", "uid": "klaus", "suspended": "0", "email": "klaus.mustermann@ceph.com"}'
```

## NFS-Ganesha Management

The Ceph Dashboard can manage [NFS Ganesha](#) exports that use CephFS or RGW as their backstore.

To enable this feature in Ceph Dashboard there are some assumptions that need to be met regarding the way NFS-Ganesha services are configured.

The dashboard manages NFS-Ganesha config files stored in RADOS objects on the Ceph Cluster. NFS-Ganesha must store part of their configuration in the Ceph cluster.

These configuration files follow the below conventions. Each export block must be stored in its own RADOS object named `export-<id>`, where `<id>` must match the `Export_ID` attribute of the export configuration. Then, for each NFS-Ganesha service daemon there should exist a RADOS object named `conf-<daemon_id>`, where `<daemon_id>` is an arbitrary string that should uniquely identify the daemon instance (e.g., the hostname where the daemon is running). Each `conf-<daemon_id>` object contains the RADOS URLs to the exports that the NFS-Ganesha daemon should serve. These URLs are of the form:

```
1. %url rados://<pool_name>[/<namespace>]/export-<id>
```

Both the `conf-<daemon_id>` and `export-<id>` objects must be stored in the same RADOS pool/namespace.

## Configuring NFS-Ganesha in the Dashboard

To enable management of NFS-Ganesha exports in the Ceph Dashboard, we need to tell the Dashboard the RADOS pool and namespace in which configuration objects are stored. The Ceph Dashboard can then access them by following the naming convention described above.

The Dashboard command to configure the NFS-Ganesha configuration objects location is:

```
1. $ ceph dashboard set-ganesha-clusters-rados-pool-namespace <pool_name>[/<namespace>]
```

After running the above command, the Ceph Dashboard is able to find the NFS-Ganesha configuration objects and we can manage exports through the Web UI.

#### Note

A dedicated pool for the NFS shares should be used. Otherwise it can cause the [known issue](#) with listing of shares if the NFS objects are stored together with a lot of other objects in a single pool.

## Support for Multiple NFS-Ganesha Clusters

The Ceph Dashboard also supports management of NFS-Ganesha exports belonging to other NFS-Ganesha clusters. An NFS-Ganesha cluster is a group of NFS-Ganesha service daemons sharing the same exports. NFS-Ganesha clusters are independent and don't share the exports configuration among each other.

Each NFS-Ganesha cluster should store its configuration objects in a unique RADOS pool/namespace to isolate the configuration.

To specify the the configuration location of each NFS-Ganesha cluster we can use the same command as above but with a different value pattern:

```
$ ceph dashboard set-ganesha-clusters-rados-pool-namespace <cluster_id>:<pool_name>[/<namespace>](),
1. <cluster_id>:<pool_name>[/<namespace>])*
```

The `<cluster_id>` is an arbitrary string that should uniquely identify the NFS-Ganesha cluster.

When configuring the Ceph Dashboard with multiple NFS-Ganesha clusters, the Web UI will allow you to choose to which cluster an export belongs.

## Support for NFS-Ganesha Clusters Deployed by the Orchestrator

The Ceph Dashboard can be used to manage NFS-Ganesha clusters deployed by the Orchestrator and will detect them automatically. For more details on deploying NFS-Ganesha clusters with the Orchestrator, please see [Stateless services \(MDS/RGW/NFS/rbd-mirror/iSCSI\)](#). Or particularly, see [Deploying NFS ganesha](#) for how to deploy NFS-Ganesha clusters with the Cephadm backend.

## Plug-ins

Plug-ins extend the functionality of the Ceph Dashboard in a modular and loosely coupled fashion.

# Feature Toggles

This plug-in allows to enable or disable some features from the Ceph Dashboard on-demand. When a feature becomes disabled:

- Its front-end elements (web pages, menu entries, charts, etc.) will become hidden.
- Its associated REST API endpoints will reject any further requests (404, Not Found Error).

The main purpose of this plug-in is to allow ad-hoc customizations of the workflows exposed by the dashboard. Additionally, it could allow for dynamically enabling experimental features with minimal configuration burden and no service impact.

The list of features that can be enabled/disabled is:

- **Block (RBD):**
  - Image Management: `rbd`
  - Mirroring: `mirroring`
  - iSCSI: `iscsi`
- **Filesystem (Cephfs):** `cephfs`
- **Objects (RGW):** `rgw` (including daemon, user and bucket management).
- **NFS:** `nfs-ganesha` exports.

By default all features come enabled.

To retrieve a list of features and their current statuses:

```

1. $ ceph dashboard feature status
2. Feature 'cephfs': 'enabled'
3. Feature 'iscsi': 'enabled'
4. Feature 'mirroring': 'enabled'
5. Feature 'rbd': 'enabled'
6. Feature 'rgw': 'enabled'
7. Feature 'nfs': 'enabled'
```

To enable or disable the status of a single or multiple features:

```

1. $ ceph dashboard feature disable iscsi mirroring
2. Feature 'iscsi': disabled
3. Feature 'mirroring': disabled
```

After a feature status has changed, the API REST endpoints immediately respond to that change, while for the front-end UI elements, it may take up to 20 seconds to reflect

it.

## Debug

This plugin allows to customize the behaviour of the dashboard according to the debug mode. It can be enabled, disabled or checked with the following command:

```
1. $ ceph dashboard debug status
2. Debug: 'disabled'
3. $ ceph dashboard debug enable
4. Debug: 'enabled'
5. $ ceph dashboard debug disable
6. Debug: 'disabled'
```

By default, it's disabled. This is the recommended setting for production deployments. If required, debug mode can be enabled without need of restarting. Currently, disabled debug mode equals to CherryPy `production` environment, while when enabled, it uses `test_suite` defaults (please refer to [CherryPy Environments](#) for more details).

It also adds request uuid (`unique_id`) to Cherrypy on versions that don't support this. It additionally prints the `unique_id` to error responses and log messages.

## Troubleshooting the Dashboard

### Locating the Dashboard

If you are unsure of the location of the Ceph Dashboard, run the following command:

```
1. $ ceph mgr services | jq .dashboard
2. "https://host:port"
```

The command returns the URL where the Ceph Dashboard is located: `https://<host>:<port>/`

#### Note

Many Ceph tools return results in JSON format. We suggest that you install the `jq` command-line utility to facilitate working with JSON data.

### Accessing the Dashboard

If you are unable to access the Ceph Dashboard, run the following commands:

1. Verify the Ceph Dashboard module is enabled:

```
1. $ ceph mgr module ls | jq .enabled_modules
```

Ensure the Ceph Dashboard module is listed in the return value of the command.  
Example snipped output from the command above:

```
1. [
2.   "dashboard",
3.   "iostat",
4.   "restful"
5. ]
```

2. If it is not listed, activate the module with the following command:

```
1. $ ceph mgr module enable dashboard
```

3. Check the Ceph Dashboard and/or `ceph-mgr` log files for any errors.

- Check if `ceph-mgr` log messages are written to a file by:

```
1. $ ceph config get mgr log_to_file
2. true
```

- Get the location of the log file (it's `/var/log/ceph/<cluster-name>-<daemon-name>.log` by default):

```
1. $ ceph config get mgr log_file
2. /var/log/ceph/$cluster-$name.log
```

4. Ensure the SSL/TSL support is configured properly:

- Check if the SSL/TSL support is enabled:

```
1. $ ceph config get mgr mgr/dashboard/ssl
```

- If the command returns `true`, verify a certificate exists by:

```
1. $ ceph config-key get mgr/dashboard/crt
```

and:

```
1. $ ceph config-key get mgr/dashboard/key
```

- If it doesn't return `true`, run the following command to generate a self-signed certificate or follow the instructions outlined in [SSL/TLS Support](#):

```
1. $ ceph dashboard create-self-signed-cert
```

## Trouble Logging into the Dashboard

If you are unable to log into the Ceph Dashboard and you receive the following error, run through the procedural checks below:



1. Check that your user credentials are correct. If you are seeing the notification message above when trying to log into the Ceph Dashboard, it is likely you are using the wrong credentials. Double check your username and password, and ensure that your keyboard's caps lock is not enabled by accident.
2. If your user credentials are correct, but you are experiencing the same error, check that the user account exists:

```
1. $ ceph dashboard ac-user-show <username>
```

This command returns your user data. If the user does not exist, it will print:

```
1. $ Error ENOENT: User <username> does not exist
```

3. Check if the user is enabled:

```
1. $ ceph dashboard ac-user-show <username> | jq .enabled  
2. true
```

Check if `enabled` is set to `true` for your user. If not the user is not enabled, run:

```
1. $ ceph dashboard ac-user-enable <username>
```

Please see [User and Role Management](#) for more information.

## A Dashboard Feature is Not Working

When an error occurs on the backend, you will usually receive an error notification on the frontend. Run through the following scenarios to debug.

1. Check the Ceph Dashboard and `ceph-mgr` logfile(s) for any errors. These can be found by searching for keywords, such as *500 Internal Server Error*, followed by `traceback`. The end of a traceback contains more details about what exact error occurred.

2. Check your web browser's Javascript Console for any errors.

## Ceph Dashboard Logs

### Dashboard Debug Flag

With this flag enabled, error traceback is included in backend responses.

To enable this flag via the Ceph Dashboard, navigate from *Cluster* to *Manager modules*. Select *Dashboard module* and click the edit button. Click the *debug* checkbox and update.

To enable it via the CLI, run the following command:

```
1. $ ceph dashboard debug enable
```

### Setting Logging Level of Dashboard Module

Setting the logging level to debug makes the log more verbose and helpful for debugging.

1. Increase the logging level of manager daemons:

```
1. $ ceph tell mgr config set debug_mgr 20
```

2. Adjust the logging level of the Ceph Dashboard module via the Dashboard or CLI:

- Navigate from *Cluster* to *Manager modules*. Select *Dashboard module* and click the edit button. Modify the `log_level` configuration.
- To adjust it via the CLI, run the following command:

```
1. $ bin/ceph config set mgr mgr/dashboard/log_level debug
```

• High log levels can result in considerable log volume, which can easily fill up your filesystem. Set a calendar reminder for an hour, a day, or a week in the future to revert this temporary logging increase. This looks something like this:

```
1. $ ceph config log  
2. ...
```

```
3. --- 11 --- 2020-11-07 11:11:11.960659 --- mgr.x/dashboard/log_level = debug ---
4. ...
5. $ ceph config reset 11
```

# API Documentation

---

## Ceph RESTful API

---

See [Ceph REST API](#).

## Ceph Storage Cluster APIs

---

See [Ceph Storage Cluster APIs](#).

## Ceph File System APIs

---

See [libcephfs](#)

## Ceph Block Device APIs

---

See [librbdpy](#).

## Ceph RADOS Gateway APIs

---

See [librgw-py](#).

## Ceph Object Store APIs

---

- See [S3-compatible API](#).
- See [Swift-compatible API](#).
- See [Admin Ops API](#).

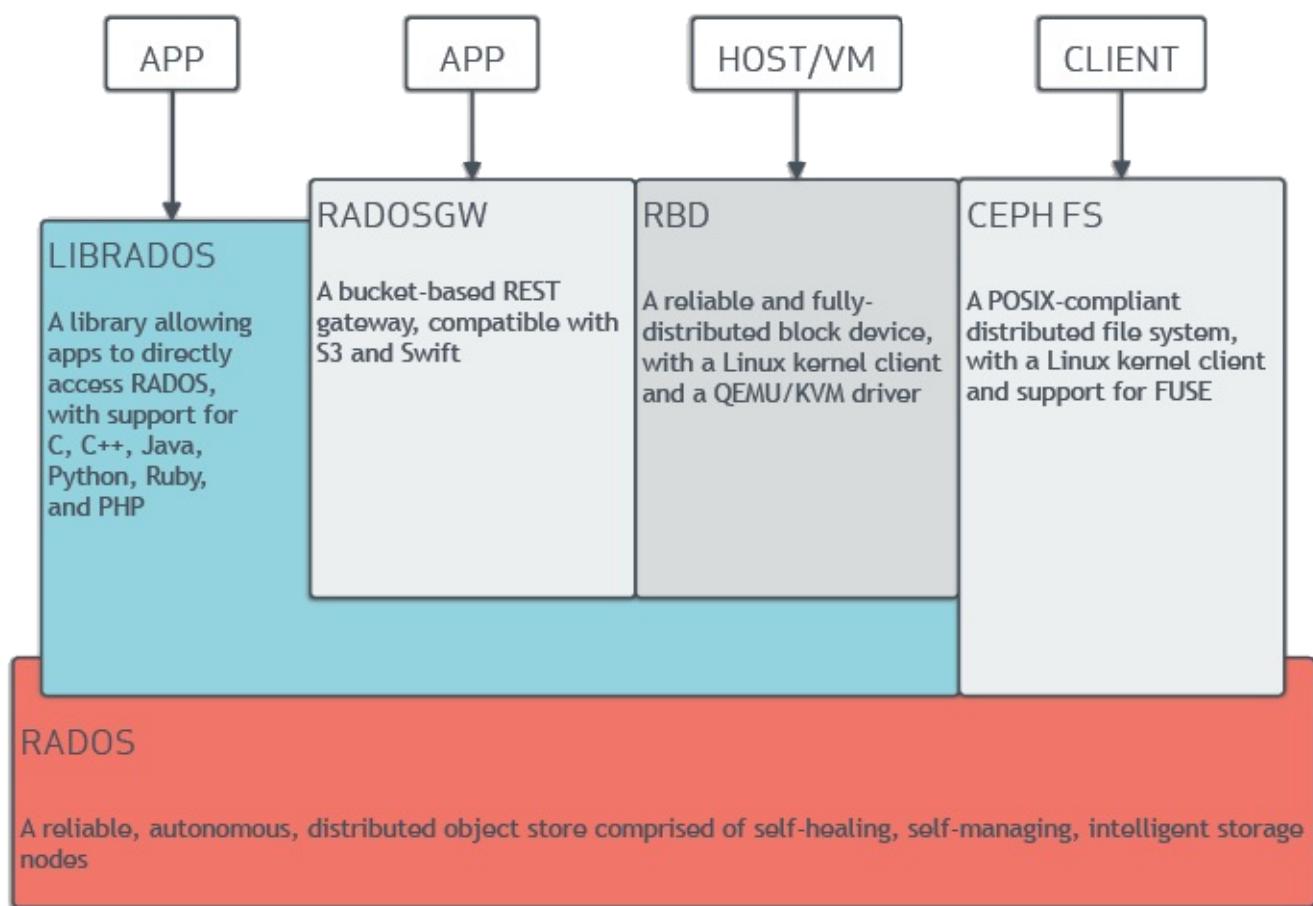
## Ceph MON Command API

---

- See [Mon command API](#).

# Architecture

Ceph uniquely delivers **object**, **block**, and **file storage** in one unified system. Ceph is highly reliable, easy to manage, and free. The power of Ceph can transform your company's IT infrastructure and your ability to manage vast amounts of data. Ceph delivers extraordinary scalability-thousands of clients accessing petabytes to exabytes of data. A **Ceph Node** leverages commodity hardware and intelligent daemons, and a **Ceph Storage Cluster** accommodates large numbers of nodes, which communicate with each other to replicate and redistribute data dynamically.



## The Ceph Storage Cluster

Ceph provides an infinitely scalable **Ceph Storage Cluster** based upon RADOS, which you can read about in [RADOS - A Scalable, Reliable Storage Service for Petabyte-scale Storage Clusters](#).

A Ceph Storage Cluster consists of multiple types of daemons:

- [Ceph Monitor](#)
- [Ceph OSD Daemon](#)

- Ceph Manager
- Ceph Metadata Server



A Ceph Monitor maintains a master copy of the cluster map. A cluster of Ceph monitors ensures high availability should a monitor daemon fail. Storage cluster clients retrieve a copy of the cluster map from the Ceph Monitor.

A Ceph OSD Daemon checks its own state and the state of other OSDs and reports back to monitors.

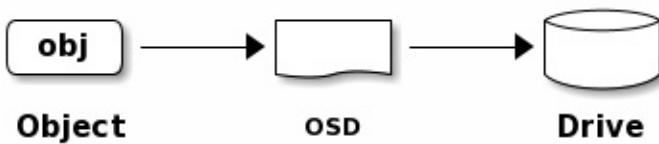
A Ceph Manager acts as an endpoint for monitoring, orchestration, and plug-in modules.

A Ceph Metadata Server (MDS) manages file metadata when CephFS is used to provide file services.

Storage cluster clients and each [Ceph OSD Daemon](#) use the CRUSH algorithm to efficiently compute information about data location, instead of having to depend on a central lookup table. Ceph's high-level features include a native interface to the Ceph Storage Cluster via [librados](#), and a number of service interfaces built on top of [librados](#).

## Storing Data

The Ceph Storage Cluster receives data from [Ceph Clients](#)—whether it comes through a [Ceph Block Device](#), [Ceph Object Storage](#), the [Ceph File System](#) or a custom implementation you create using [librados](#)—which is stored as RADOS objects. Each object is stored on an [Object Storage Device](#). Ceph OSD Daemons handle read, write, and replication operations on storage drives. With the older Filestore back end, each RADOS object was stored as a separate file on a conventional filesystem (usually XFS). With the new and default BlueStore back end, objects are stored in a monolithic database-like fashion.



Ceph OSD Daemons store data as objects in a flat namespace (e.g., no hierarchy of directories). An object has an identifier, binary data, and metadata consisting of a set of name/value pairs. The semantics are completely up to [Ceph Clients](#). For example,

CephFS uses metadata to store file attributes such as the file owner, created date, last modified date, and so forth.

ID	Binary Data	Metadata
1234	01010101010100110101010010 0101100001010100110101010010 0101100001010100110101010010	name1 value1 name2 value2 nameN valueN

#### Note

An object ID is unique across the entire cluster, not just the local filesystem.

## Scalability and High Availability

In traditional architectures, clients talk to a centralized component (e.g., a gateway, broker, API, facade, etc.), which acts as a single point of entry to a complex subsystem. This imposes a limit to both performance and scalability, while introducing a single point of failure (i.e., if the centralized component goes down, the whole system goes down, too).

Ceph eliminates the centralized gateway to enable clients to interact with Ceph OSD Daemons directly. Ceph OSD Daemons create object replicas on other Ceph Nodes to ensure data safety and high availability. Ceph also uses a cluster of monitors to ensure high availability. To eliminate centralization, Ceph uses an algorithm called CRUSH.

## CRUSH Introduction

Ceph Clients and Ceph OSD Daemons both use the CRUSH algorithm to efficiently compute information about object location, instead of having to depend on a central lookup table. CRUSH provides a better data management mechanism compared to older approaches, and enables massive scale by cleanly distributing the work to all the clients and OSD daemons in the cluster. CRUSH uses intelligent data replication to ensure resiliency, which is better suited to hyper-scale storage. The following sections provide additional details on how CRUSH works. For a detailed discussion of CRUSH, see [CRUSH - Controlled, Scalable, Decentralized Placement of Replicated Data](#).

## Cluster Map

Ceph depends upon Ceph Clients and Ceph OSD Daemons having knowledge of the cluster topology, which is inclusive of 5 maps collectively referred to as the “Cluster Map”:

1. **The Monitor Map:** Contains the cluster `fsid`, the position, name address and port of each monitor. It also indicates the current epoch, when the map was created, and the last time it changed. To view a monitor map, execute `ceph mon dump`.

2. **The OSD Map:** Contains the cluster `fsid`, when the map was created and last modified, a list of pools, replica sizes, PG numbers, a list of OSDs and their status (e.g., `up`, `in`). To view an OSD map, execute `ceph osd dump`.
3. **The PG Map:** Contains the PG version, its time stamp, the last OSD map epoch, the full ratios, and details on each placement group such as the PG ID, the Up Set, the Acting Set, the state of the PG (e.g., `active + clean`), and data usage statistics for each pool.
4. **The CRUSH Map:** Contains a list of storage devices, the failure domain hierarchy (e.g., device, host, rack, row, room, etc.), and rules for traversing the hierarchy when storing data. To view a CRUSH map, execute `ceph osd getcrushmap -o {filename}`; then, decompile it by executing `crushtool -d {comp-crushmap-filename} -o {decomp-crushmap-filename}`. You can view the decompiled map in a text editor or with `cat`.
5. **The MDS Map:** Contains the current MDS map epoch, when the map was created, and the last time it changed. It also contains the pool for storing metadata, a list of metadata servers, and which metadata servers are `up` and `in`. To view an MDS map, execute `ceph fs dump`.

Each map maintains an iterative history of its operating state changes. Ceph Monitors maintain a master copy of the cluster map including the cluster members, state, changes, and the overall health of the Ceph Storage Cluster.

## High Availability Monitors

Before Ceph Clients can read or write data, they must contact a Ceph Monitor to obtain the most recent copy of the cluster map. A Ceph Storage Cluster can operate with a single monitor; however, this introduces a single point of failure (i.e., if the monitor goes down, Ceph Clients cannot read or write data).

For added reliability and fault tolerance, Ceph supports a cluster of monitors. In a cluster of monitors, latency and other faults can cause one or more monitors to fall behind the current state of the cluster. For this reason, Ceph must have agreement among various monitor instances regarding the state of the cluster. Ceph always uses a majority of monitors (e.g., 1, 2:3, 3:5, 4:6, etc.) and the [Paxos](#) algorithm to establish a consensus among the monitors about the current state of the cluster.

For details on configuring monitors, see the [Monitor Config Reference](#).

## High Availability Authentication

To identify users and protect against man-in-the-middle attacks, Ceph provides its `cephx` authentication system to authenticate users and daemons.

### Note

The `cephx` protocol does not address data encryption in transport (e.g., SSL/TLS) or encryption at rest.

Ceph uses shared secret keys for authentication, meaning both the client and the monitor cluster have a copy of the client's secret key. The authentication protocol is such that both parties are able to prove to each other they have a copy of the key without actually revealing it. This provides mutual authentication, which means the cluster is sure the user possesses the secret key, and the user is sure that the cluster has a copy of the secret key.

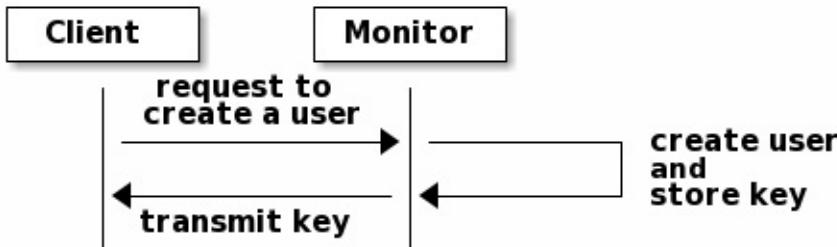
A key scalability feature of Ceph is to avoid a centralized interface to the Ceph object store, which means that Ceph clients must be able to interact with OSDs directly. To protect data, Ceph provides its `cephx` authentication system, which authenticates users operating Ceph clients. The `cephx` protocol operates in a manner with behavior similar to [Kerberos](#).

A user/actor invokes a Ceph client to contact a monitor. Unlike Kerberos, each monitor can authenticate users and distribute keys, so there is no single point of failure or bottleneck when using `cephx`. The monitor returns an authentication data structure similar to a Kerberos ticket that contains a session key for use in obtaining Ceph services. This session key is itself encrypted with the user's permanent secret key, so that only the user can request services from the Ceph Monitor(s). The client then uses the session key to request its desired services from the monitor, and the monitor provides the client with a ticket that will authenticate the client to the OSDs that actually handle data. Ceph Monitors and OSDs share a secret, so the client can use the ticket provided by the monitor with any OSD or metadata server in the cluster. Like Kerberos, `cephx` tickets expire, so an attacker cannot use an expired ticket or session key obtained surreptitiously. This form of authentication will prevent attackers with access to the communications medium from either creating bogus messages under another user's identity or altering another user's legitimate messages, as long as the user's secret key is not divulged before it expires.

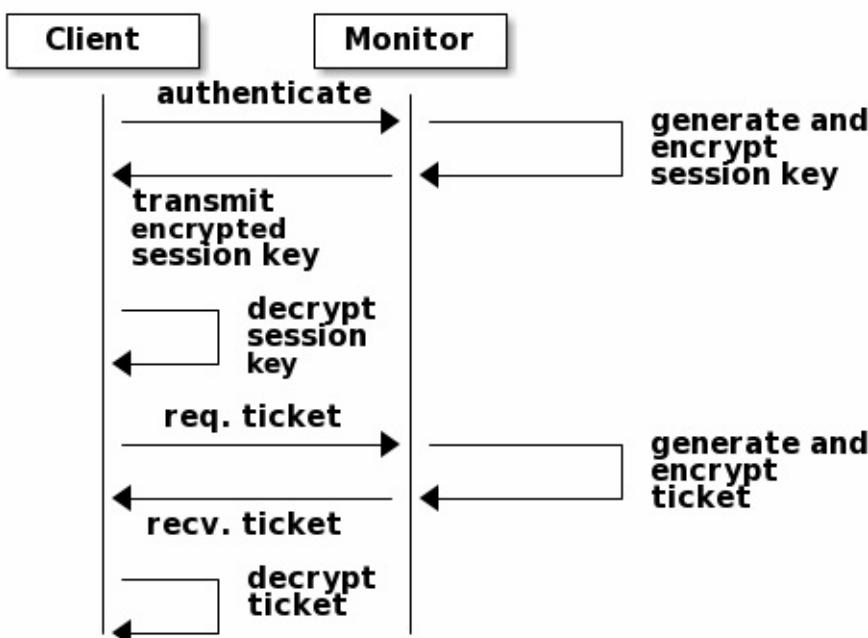
To use `cephx`, an administrator must set up users first. In the following diagram, the `client.admin` user invokes `ceph auth get-or-create-key` from the command line to generate a username and secret key. Ceph's `auth` subsystem generates the username and key, stores a copy with the monitor(s) and transmits the user's secret back to the `client.admin` user. This means that the client and the monitor share a secret key.

#### Note

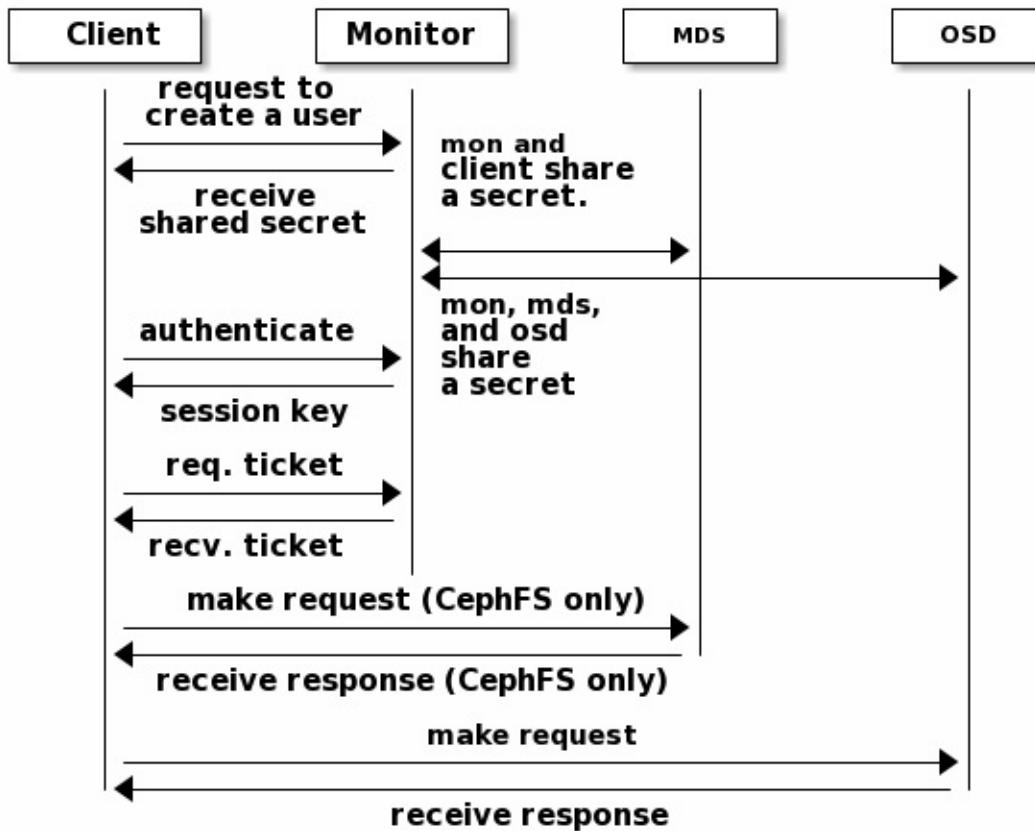
The `client.admin` user must provide the user ID and secret key to the user in a secure manner.



To authenticate with the monitor, the client passes in the user name to the monitor, and the monitor generates a session key and encrypts it with the secret key associated to the user name. Then, the monitor transmits the encrypted ticket back to the client. The client then decrypts the payload with the shared secret key to retrieve the session key. The session key identifies the user for the current session. The client then requests a ticket on behalf of the user signed by the session key. The monitor generates a ticket, encrypts it with the user's secret key and transmits it back to the client. The client decrypts the ticket and uses it to sign requests to OSDs and metadata servers throughout the cluster.



The `cephx` protocol authenticates ongoing communications between the client machine and the Ceph servers. Each message sent between a client and server, subsequent to the initial authentication, is signed using a ticket that the monitors, OSDs and metadata servers can verify with their shared secret.



The protection offered by this authentication is between the Ceph client and the Ceph server hosts. The authentication is not extended beyond the Ceph client. If the user accesses the Ceph client from a remote host, Ceph authentication is not applied to the connection between the user's host and the client host.

For configuration details, see [Cephx Config Guide](#). For user management details, see [User Management](#).

## Smart Daemons Enable Hyperscale

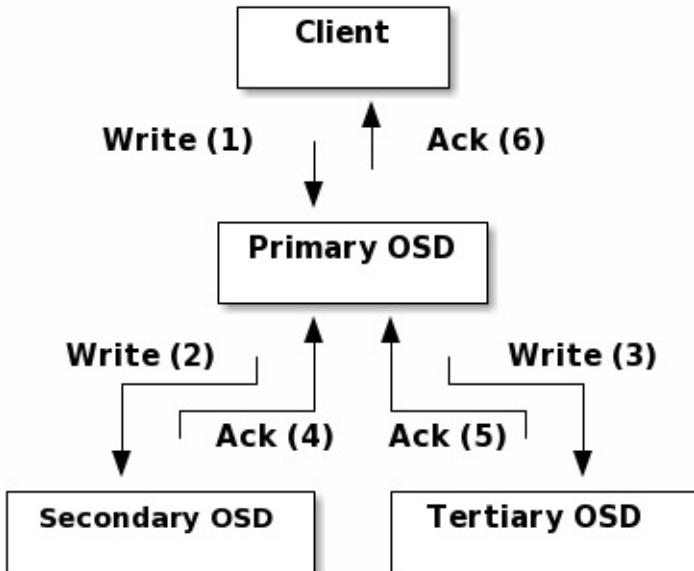
In many clustered architectures, the primary purpose of cluster membership is so that a centralized interface knows which nodes it can access. Then the centralized interface provides services to the client through a double dispatch—which is a **huge** bottleneck at the petabyte-to-exabyte scale.

Ceph eliminates the bottleneck: Ceph's OSD Daemons AND Ceph Clients are cluster aware. Like Ceph clients, each Ceph OSD Daemon knows about other Ceph OSD Daemons in the cluster. This enables Ceph OSD Daemons to interact directly with other Ceph OSD Daemons and Ceph Monitors. Additionally, it enables Ceph Clients to interact directly with Ceph OSD Daemons.

The ability of Ceph Clients, Ceph Monitors and Ceph OSD Daemons to interact with each other means that Ceph OSD Daemons can utilize the CPU and RAM of the Ceph nodes to easily perform tasks that would bog down a centralized server. The ability to leverage this computing power leads to several major benefits:

1. **OSDs Service Clients Directly:** Since any network device has a limit to the number of concurrent connections it can support, a centralized system has a low physical limit at high scales. By enabling Ceph Clients to contact Ceph OSD Daemons directly, Ceph increases both performance and total system capacity simultaneously, while removing a single point of failure. Ceph Clients can maintain a session when they need to, and with a particular Ceph OSD Daemon instead of a centralized server.
2. **OSD Membership and Status:** Ceph OSD Daemons join a cluster and report on their status. At the lowest level, the Ceph OSD Daemon status is `up` or `down` reflecting whether or not it is running and able to service Ceph Client requests. If a Ceph OSD Daemon is `down` and `in` the Ceph Storage Cluster, this status may indicate the failure of the Ceph OSD Daemon. If a Ceph OSD Daemon is not running (e.g., it crashes), the Ceph OSD Daemon cannot notify the Ceph Monitor that it is `down`. The OSDs periodically send messages to the Ceph Monitor (`MPGStats` pre-luminous, and a new `MOSDBeacon` in luminous). If the Ceph Monitor doesn't see that message after a configurable period of time then it marks the OSD down. This mechanism is a failsafe, however. Normally, Ceph OSD Daemons will determine if a neighboring OSD is down and report it to the Ceph Monitor(s). This assures that Ceph Monitors are lightweight processes. See [Monitoring OSDs](#) and [Heartbeats](#) for additional details.
3. **Data Scrubbing:** As part of maintaining data consistency and cleanliness, Ceph OSD Daemons can scrub objects. That is, Ceph OSD Daemons can compare their local objects metadata with its replicas stored on other OSDs. Scrubbing happens on a per-Placement Group base. Scrubbing (usually performed daily) catches mismatches in size and other metadata. Ceph OSD Daemons also perform deeper scrubbing by comparing data in objects bit-for-bit with their checksums. Deep scrubbing (usually performed weekly) finds bad sectors on a drive that weren't apparent in a light scrub. See [Data Scrubbing](#) for details on configuring scrubbing.
4. **Replication:** Like Ceph Clients, Ceph OSD Daemons use the CRUSH algorithm, but the Ceph OSD Daemon uses it to compute where replicas of objects should be stored (and for rebalancing). In a typical write scenario, a client uses the CRUSH algorithm to compute where to store an object, maps the object to a pool and placement group, then looks at the CRUSH map to identify the primary OSD for the placement group.

The client writes the object to the identified placement group in the primary OSD. Then, the primary OSD with its own copy of the CRUSH map identifies the secondary and tertiary OSDs for replication purposes, and replicates the object to the appropriate placement groups in the secondary and tertiary OSDs (as many OSDs as additional replicas), and responds to the client once it has confirmed the object was stored successfully.



With the ability to perform data replication, Ceph OSD Daemons relieve Ceph clients from that duty, while ensuring high data availability and data safety.

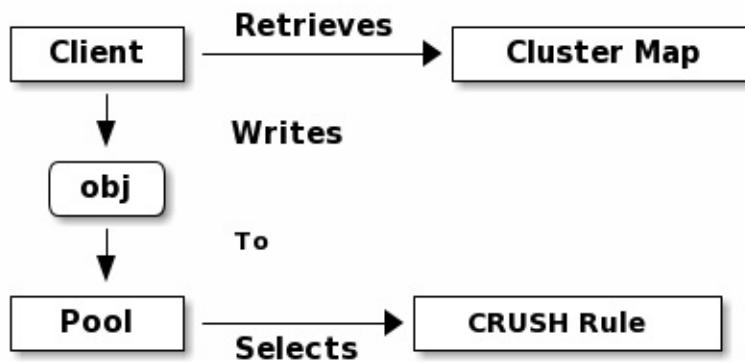
## Dynamic Cluster Management

In the [Scalability and High Availability](#) section, we explained how Ceph uses CRUSH, cluster awareness and intelligent daemons to scale and maintain high availability. Key to Ceph's design is the autonomous, self-healing, and intelligent Ceph OSD Daemon. Let's take a deeper look at how CRUSH works to enable modern cloud storage infrastructures to place data, rebalance the cluster and recover from faults dynamically.

## About Pools

The Ceph storage system supports the notion of 'Pools', which are logical partitions for storing objects.

Ceph Clients retrieve a [Cluster Map](#) from a Ceph Monitor, and write objects to pools. The pool's `size` or number of replicas, the CRUSH rule and the number of placement groups determine how Ceph will place the data.



Pools set at least the following parameters:

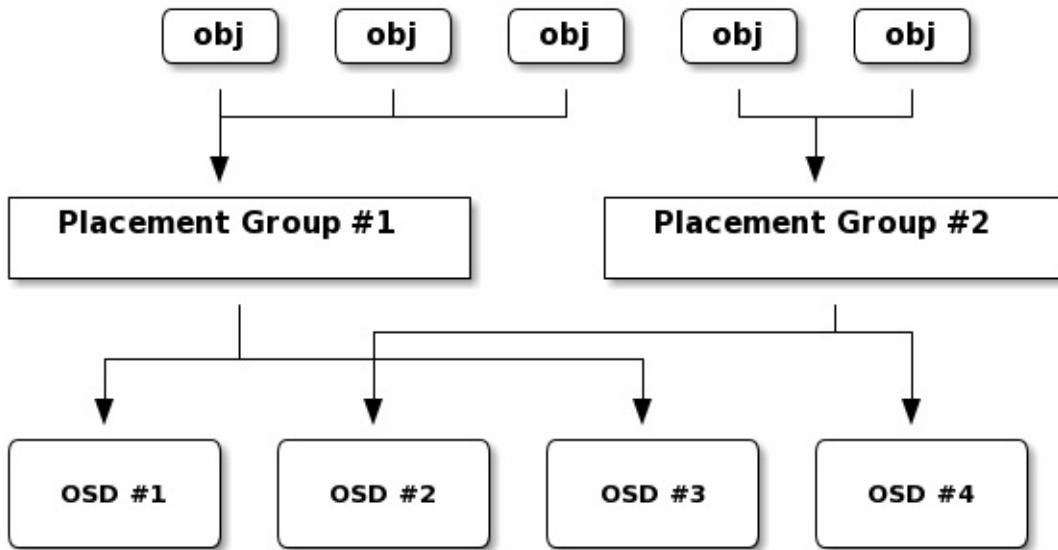
- Ownership/Access to Objects
- The Number of Placement Groups, and
- The CRUSH Rule to Use.

See [Set Pool Values](#) for details.

## Mapping PGs to OSDs

Each pool has a number of placement groups. CRUSH maps PGs to OSDs dynamically. When a Ceph Client stores objects, CRUSH will map each object to a placement group.

Mapping objects to placement groups creates a layer of indirection between the Ceph OSD Daemon and the Ceph Client. The Ceph Storage Cluster must be able to grow (or shrink) and rebalance where it stores objects dynamically. If the Ceph Client “knew” which Ceph OSD Daemon had which object, that would create a tight coupling between the Ceph Client and the Ceph OSD Daemon. Instead, the CRUSH algorithm maps each object to a placement group and then maps each placement group to one or more Ceph OSD Daemons. This layer of indirection allows Ceph to rebalance dynamically when new Ceph OSD Daemons and the underlying OSD devices come online. The following diagram depicts how CRUSH maps objects to placement groups, and placement groups to OSDs.



With a copy of the cluster map and the CRUSH algorithm, the client can compute exactly which OSD to use when reading or writing a particular object.

## Calculating PG IDs

When a Ceph Client binds to a Ceph Monitor, it retrieves the latest copy of the [Cluster Map](#). With the cluster map, the client knows about all of the monitors, OSDs, and metadata servers in the cluster. **However, it doesn't know anything about object locations.**

Object locations get computed.

The only input required by the client is the object ID and the pool. It's simple: Ceph stores data in named pools (e.g., "liverpool"). When a client wants to store a named object (e.g., "john," "paul," "george," "ringo", etc.) it calculates a placement group using the object name, a hash code, the number of PGs in the pool and the pool name. Ceph clients use the following steps to compute PG IDs.

1. The client inputs the pool name and the object ID. (e.g., pool = "liverpool" and object-id = "john")
2. Ceph takes the object ID and hashes it.
3. Ceph calculates the hash modulo the number of PGs. (e.g., [58](#)) to get a PG ID.
4. Ceph gets the pool ID given the pool name (e.g., "liverpool" = [4](#))
5. Ceph prepends the pool ID to the PG ID (e.g., [4.58](#)).

Computing object locations is much faster than performing object location query over a chatty session. The CRUSH algorithm allows a client to compute where objects *should* be stored, and enables the client to contact the primary OSD to store or retrieve the

objects.

## Peering and Sets

In previous sections, we noted that Ceph OSD Daemons check each others heartbeats and report back to the Ceph Monitor. Another thing Ceph OSD daemons do is called ‘peering’, which is the process of bringing all of the OSDs that store a Placement Group (PG) into agreement about the state of all of the objects (and their metadata) in that PG. In fact, Ceph OSD Daemons [Report Peering Failure](#) to the Ceph Monitors. Peering issues usually resolve themselves; however, if the problem persists, you may need to refer to the [Troubleshooting Peering Failure](#) section.

### Note

Agreeing on the state does not mean that the PGs have the latest contents.

The Ceph Storage Cluster was designed to store at least two copies of an object (i.e., `size = 2`), which is the minimum requirement for data safety. For high availability, a Ceph Storage Cluster should store more than two copies of an object (e.g., `size = 3` and `min size = 2`) so that it can continue to run in a `degraded` state while maintaining data safety.

Referring back to the diagram in [Smart Daemons Enable Hyperscale](#), we do not name the Ceph OSD Daemons specifically (e.g., `osd.0`, `osd.1`, etc.), but rather refer to them as *Primary*, *Secondary*, and so forth. By convention, the *Primary* is the first OSD in the *Acting Set*, and is responsible for coordinating the peering process for each placement group where it acts as the *Primary*, and is the **ONLY** OSD that will accept client-initiated writes to objects for a given placement group where it acts as the *Primary*.

When a series of OSDs are responsible for a placement group, that series of OSDs, we refer to them as an *Acting Set*. An *Acting Set* may refer to the Ceph OSD Daemons that are currently responsible for the placement group, or the Ceph OSD Daemons that were responsible for a particular placement group as of some epoch.

The Ceph OSD daemons that are part of an *Acting Set* may not always be `up`. When an OSD in the *Acting Set* is `up`, it is part of the *Up Set*. The *Up Set* is an important distinction, because Ceph can remap PGs to other Ceph OSD Daemons when an OSD fails.

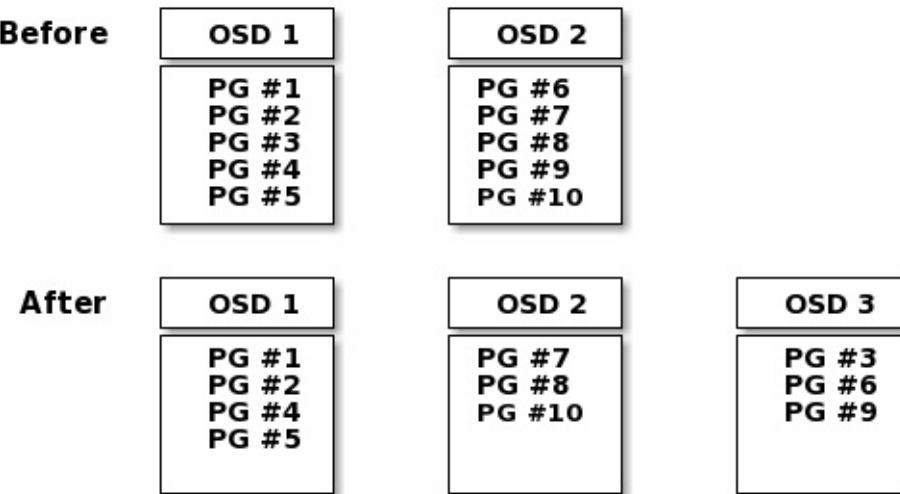
### Note

In an *Acting Set* for a PG containing `osd.25`, `osd.32` and `osd.61`, the first OSD, `osd.25`, is the *Primary*. If that OSD fails, the Secondary, `osd.32`, becomes the *Primary*, and `osd.25` will be removed from the *Up Set*.

## Rebalancing

When you add a Ceph OSD Daemon to a Ceph Storage Cluster, the cluster map gets updated with the new OSD. Referring back to [Calculating PG IDs](#), this changes the cluster map.

Consequently, it changes object placement, because it changes an input for the calculations. The following diagram depicts the rebalancing process (albeit rather crudely, since it is substantially less impactful with large clusters) where some, but not all of the PGs migrate from existing OSDs (OSD 1, and OSD 2) to the new OSD (OSD 3). Even when rebalancing, CRUSH is stable. Many of the placement groups remain in their original configuration, and each OSD gets some added capacity, so there are no load spikes on the new OSD after rebalancing is complete.



## Data Consistency

As part of maintaining data consistency and cleanliness, Ceph OSDs also scrub objects within placement groups. That is, Ceph OSDs compare object metadata in one placement group with its replicas in placement groups stored in other OSDs. Scrubbing (usually performed daily) catches OSD bugs or filesystem errors, often as a result of hardware issues. OSDs also perform deeper scrubbing by comparing data in objects bit-for-bit. Deep scrubbing (by default performed weekly) finds bad blocks on a drive that weren't apparent in a light scrub.

See [Data Scrubbing](#) for details on configuring scrubbing.

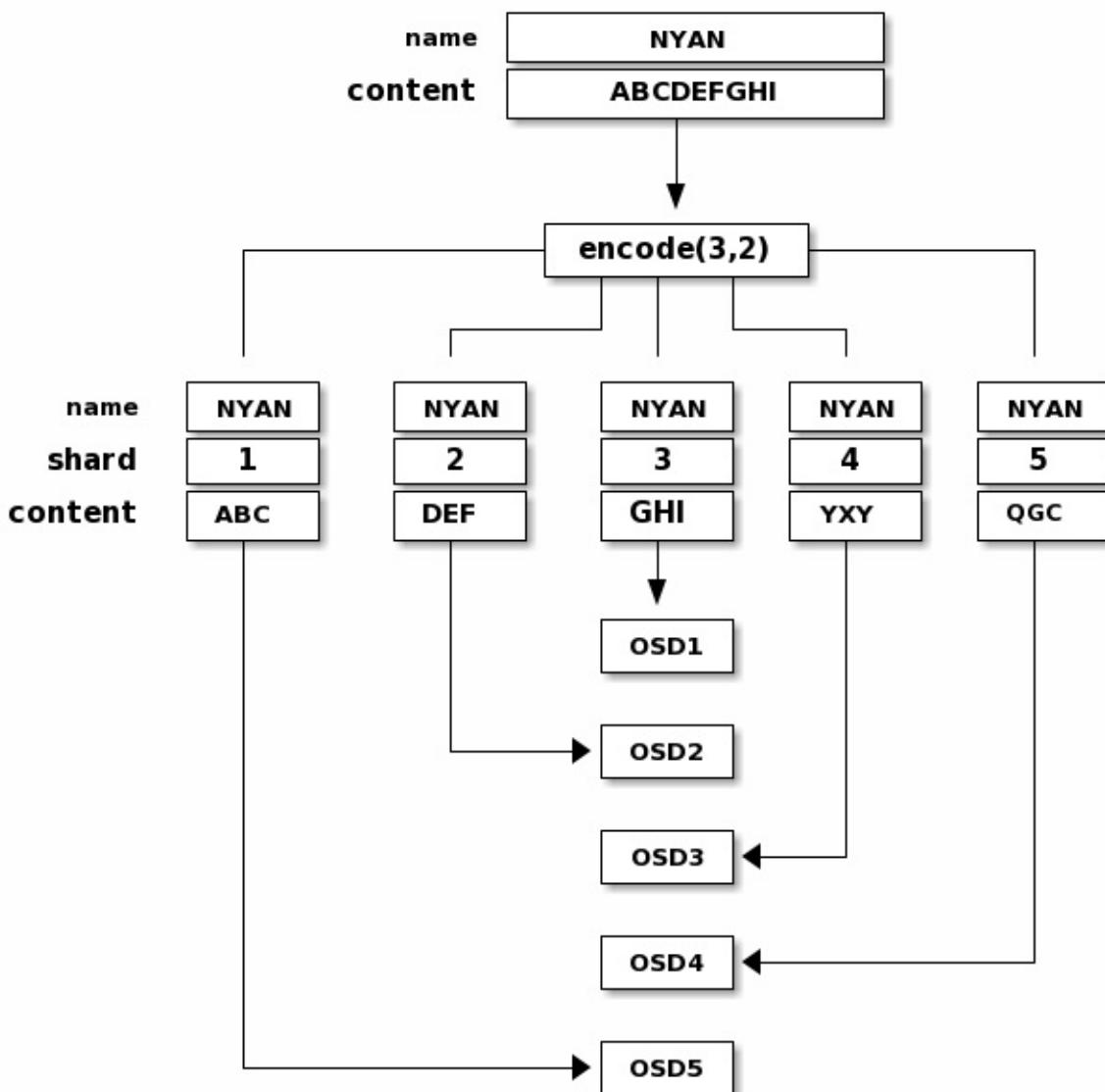
## Erasure Coding

An erasure coded pool stores each object as  $K+M$  chunks. It is divided into  $K$  data chunks and  $M$  coding chunks. The pool is configured to have a size of  $K+M$  so that each chunk is stored in an OSD in the acting set. The rank of the chunk is stored as an attribute of the object.

For instance an erasure coded pool can be created to use five OSDs ( $K+M = 5$ ) and sustain the loss of two of them ( $M = 2$ ).

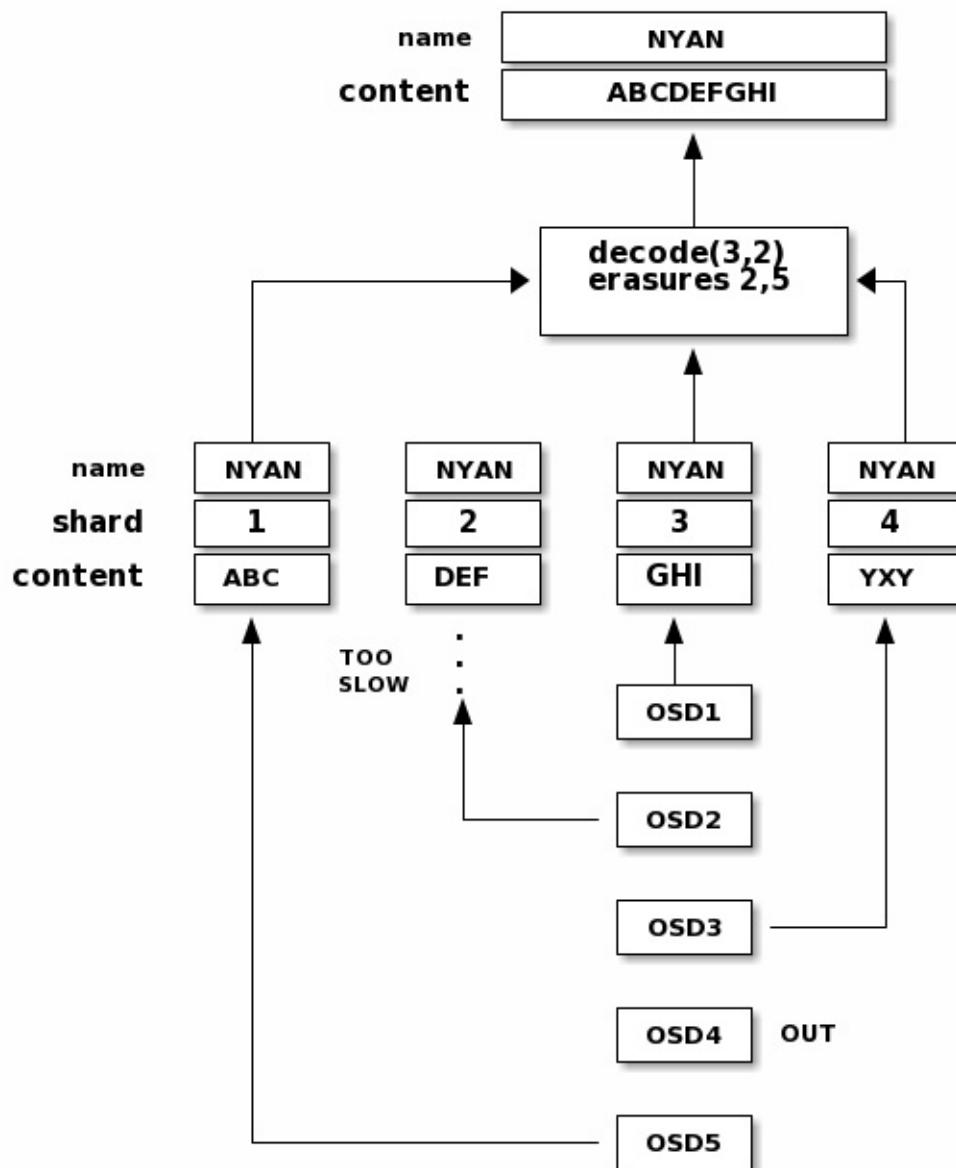
## Reading and Writing Encoded Chunks

When the object **NYAN** containing `ABCDEFGHI` is written to the pool, the erasure encoding function splits the content into three data chunks simply by dividing the content in three: the first contains `ABC`, the second `DEF` and the last `GHI`. The content will be padded if the content length is not a multiple of `K`. The function also creates two coding chunks: the fourth with `YXY` and the fifth with `QGC`. Each chunk is stored in an OSD in the acting set. The chunks are stored in objects that have the same name (**NYAN**) but reside on different OSDs. The order in which the chunks were created must be preserved and is stored as an attribute of the object (`shard_t`), in addition to its name. Chunk 1 contains `ABC` and is stored on **OSD5** while chunk 4 contains `YXY` and is stored on **OSD3**.



When the object **NYAN** is read from the erasure coded pool, the decoding function reads three chunks: chunk 1 containing `ABC`, chunk 3 containing `GHI` and chunk 4 containing `YXY`. Then, it rebuilds the original content of the object `ABCDEFGHI`. The decoding function is informed that the chunks 2 and 5 are missing (they are called 'erasures'). The chunk 5 could not be read because the **OSD4** is out. The decoding function can be called as soon as three chunks are read: **OSD2** was the slowest and its

chunk was not taken into account.

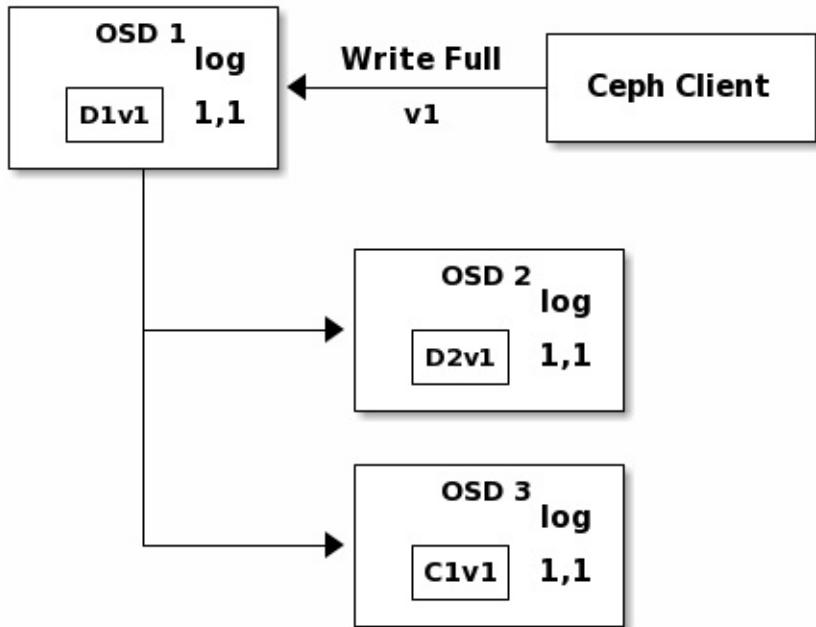


## Interrupted Full Writes

In an erasure coded pool, the primary OSD in the up set receives all write operations. It is responsible for encoding the payload into  $K+M$  chunks and sends them to the other OSDs. It is also responsible for maintaining an authoritative version of the placement group logs.

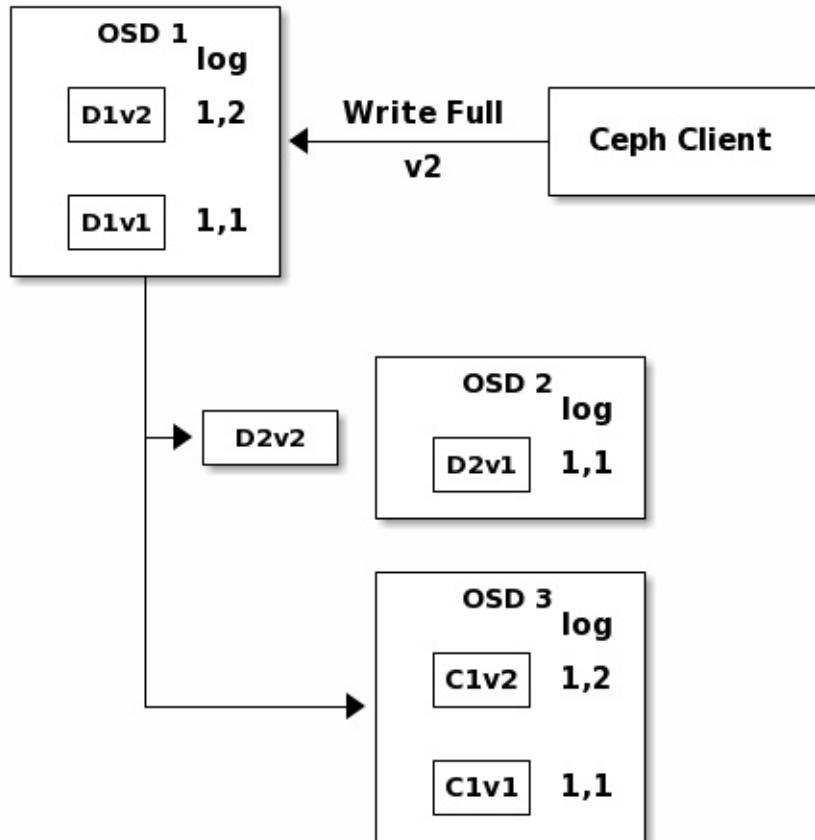
In the following diagram, an erasure coded placement group has been created with  $K = 2, M = 1$  and is supported by three OSDs, two for  $K$  and one for  $M$ . The acting set of the placement group is made of **OSD 1**, **OSD 2** and **OSD 3**. An object has been encoded and stored in the OSDs : the chunk **D1v1** (i.e. Data chunk number 1, version 1) is on **OSD 1**, **D2v1** on **OSD 2** and **C1v1** (i.e. Coding chunk number 1, version 1) on **OSD 3**. The placement group logs on each OSD are identical (i.e. **1,1** for epoch 1, version 1).

## Primary OSD



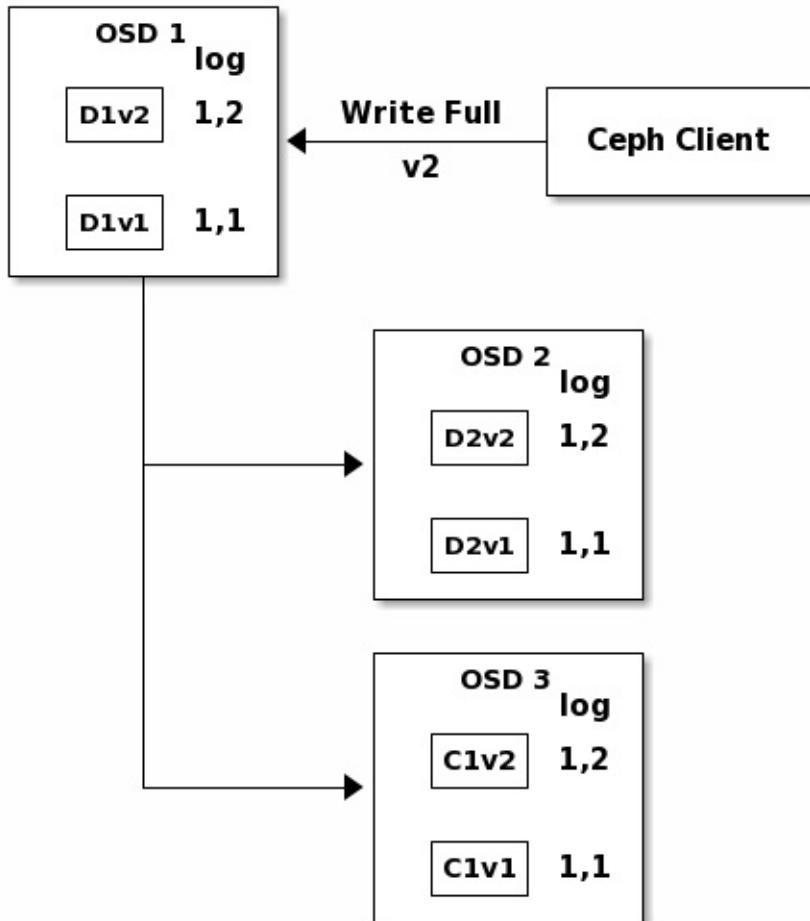
**OSD 1** is the primary and receives a **WRITE FULL** from a client, which means the payload is to replace the object entirely instead of overwriting a portion of it. Version 2 (v2) of the object is created to override version 1 (v1). **OSD 1** encodes the payload into three chunks: **D1v2** (i.e. Data chunk number 1 version 2) will be on **OSD 1**, **D2v2** on **OSD 2** and **C1v2** (i.e. Coding chunk number 1 version 2) on **OSD 3**. Each chunk is sent to the target OSD, including the primary OSD which is responsible for storing chunks in addition to handling write operations and maintaining an authoritative version of the placement group logs. When an OSD receives the message instructing it to write the chunk, it also creates a new entry in the placement group logs to reflect the change. For instance, as soon as **OSD 3** stores **C1v2**, it adds the entry **1,2** (i.e. epoch 1, version 2) to its logs. Because the OSDs work asynchronously, some chunks may still be in flight (such as **D2v2**) while others are acknowledged and persisted to storage drives (such as **C1v1** and **D1v1**).

## Primary OSD



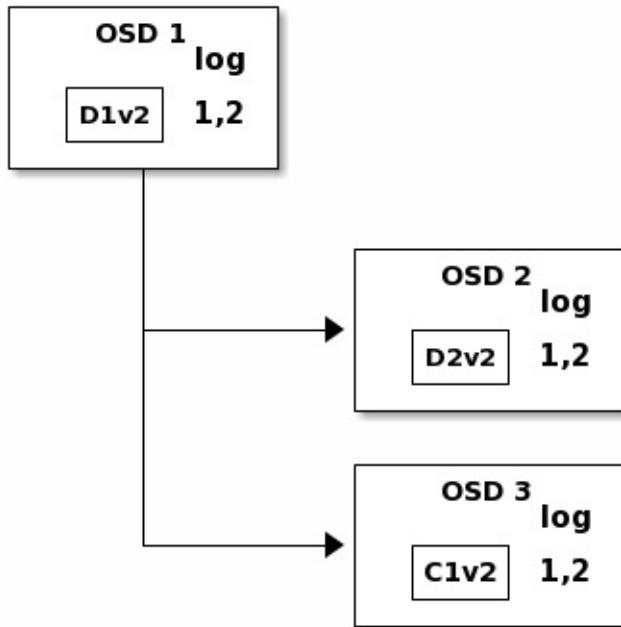
If all goes well, the chunks are acknowledged on each OSD in the acting set and the logs' `last_complete` pointer can move from `1,1` to `1,2`.

## Primary OSD

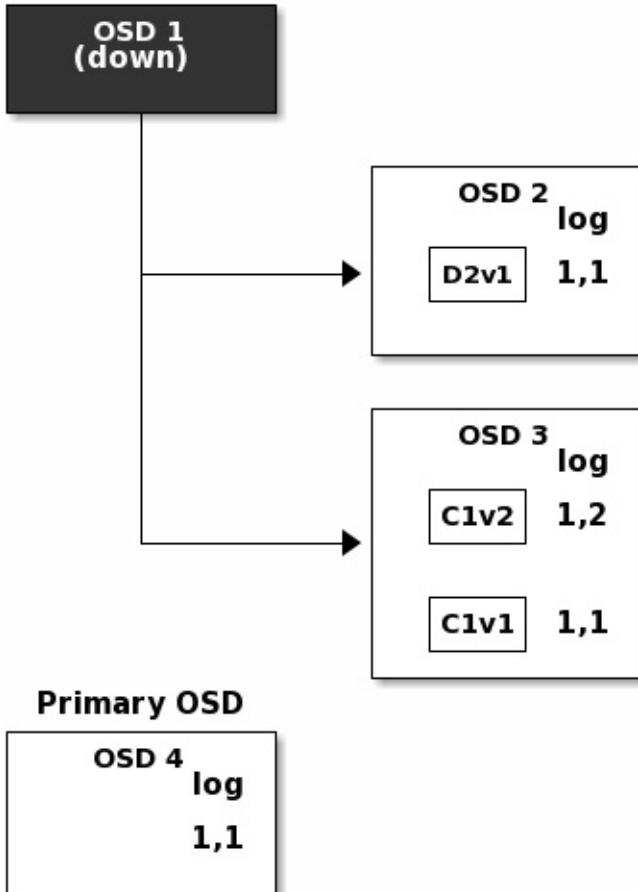


Finally, the files used to store the chunks of the previous version of the object can be removed: **D1v1** on **OSD 1**, **D2v1** on **OSD 2** and **C1v1** on **OSD 3**.

## Primary OSD

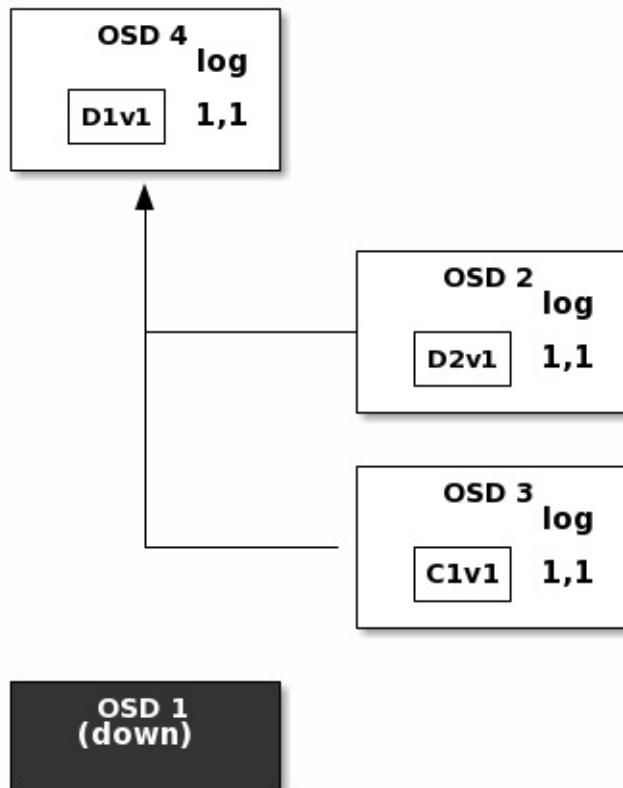


But accidents happen. If **OSD 1** goes down while **D2v2** is still in flight, the object's version 2 is partially written: **OSD 3** has one chunk but that is not enough to recover. It lost two chunks: **D1v2** and **D2v2** and the erasure coding parameters **K = 2**, **M = 1** require that at least two chunks are available to rebuild the third. **OSD 4** becomes the new primary and finds that the **last\_complete** log entry (i.e., all objects before this entry were known to be available on all OSDs in the previous acting set) is **1,1** and that will be the head of the new authoritative log.



The log entry `1,2` found on **OSD 3** is divergent from the new authoritative log provided by **OSD 4**: it is discarded and the file containing the `c1v2` chunk is removed. The `D1v1` chunk is rebuilt with the `decode` function of the erasure coding library during scrubbing and stored on the new primary **OSD 4**.

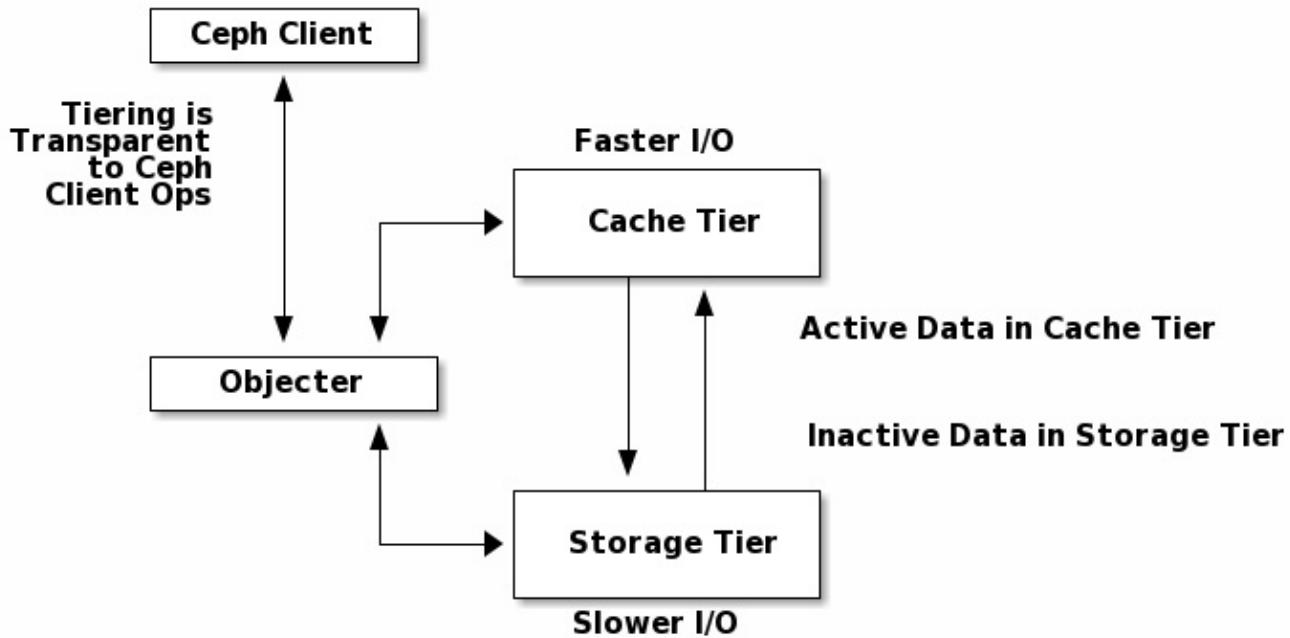
## Primary OSD



See [Erasure Code Notes](#) for additional details.

## Cache Tiering

A cache tier provides Ceph Clients with better I/O performance for a subset of the data stored in a backing storage tier. Cache tiering involves creating a pool of relatively fast/expensive storage devices (e.g., solid state drives) configured to act as a cache tier, and a backing pool of either erasure-coded or relatively slower/cheaper devices configured to act as an economical storage tier. The Ceph objecter handles where to place the objects and the tiering agent determines when to flush objects from the cache to the backing storage tier. So the cache tier and the backing storage tier are completely transparent to Ceph clients.



See [Cache Tiering](#) for additional details. Note that Cache Tiers can be tricky and their use is now discouraged.

## Extending Ceph

You can extend Ceph by creating shared object classes called ‘Ceph Classes’. Ceph loads `.so` classes stored in the `osd class dir` directory dynamically (i.e., `$libdir/rados-classes` by default). When you implement a class, you can create new object methods that have the ability to call the native methods in the Ceph Object Store, or other class methods you incorporate via libraries or create yourself.

On writes, Ceph Classes can call native or class methods, perform any series of operations on the inbound data and generate a resulting write transaction that Ceph will apply atomically.

On reads, Ceph Classes can call native or class methods, perform any series of operations on the outbound data and return the data to the client.

### Ceph Class Example

A Ceph class for a content management system that presents pictures of a particular size and aspect ratio could take an inbound bitmap image, crop it to a particular aspect ratio, resize it and embed an invisible copyright or watermark to help protect the intellectual property; then, save the resulting bitmap image to the object store.

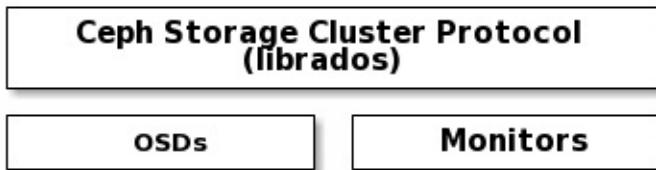
See [src/objclass/objclass.h](#) , [src/fooclass.cc](#) and [src/barclass](#) for exemplary implementations.

## Summary

Ceph Storage Clusters are dynamic-like a living organism. Whereas, many storage appliances do not fully utilize the CPU and RAM of a typical commodity server, Ceph does. From heartbeats, to peering, to rebalancing the cluster or recovering from faults, Ceph offloads work from clients (and from a centralized gateway which doesn't exist in the Ceph architecture) and uses the computing power of the OSDs to perform the work. When referring to [Hardware Recommendations](#) and the [Network Config Reference](#), be cognizant of the foregoing concepts to understand how Ceph utilizes computing resources.

## Ceph Protocol

Ceph Clients use the native protocol for interacting with the Ceph Storage Cluster. Ceph packages this functionality into the `librados` library so that you can create your own custom Ceph Clients. The following diagram depicts the basic architecture.



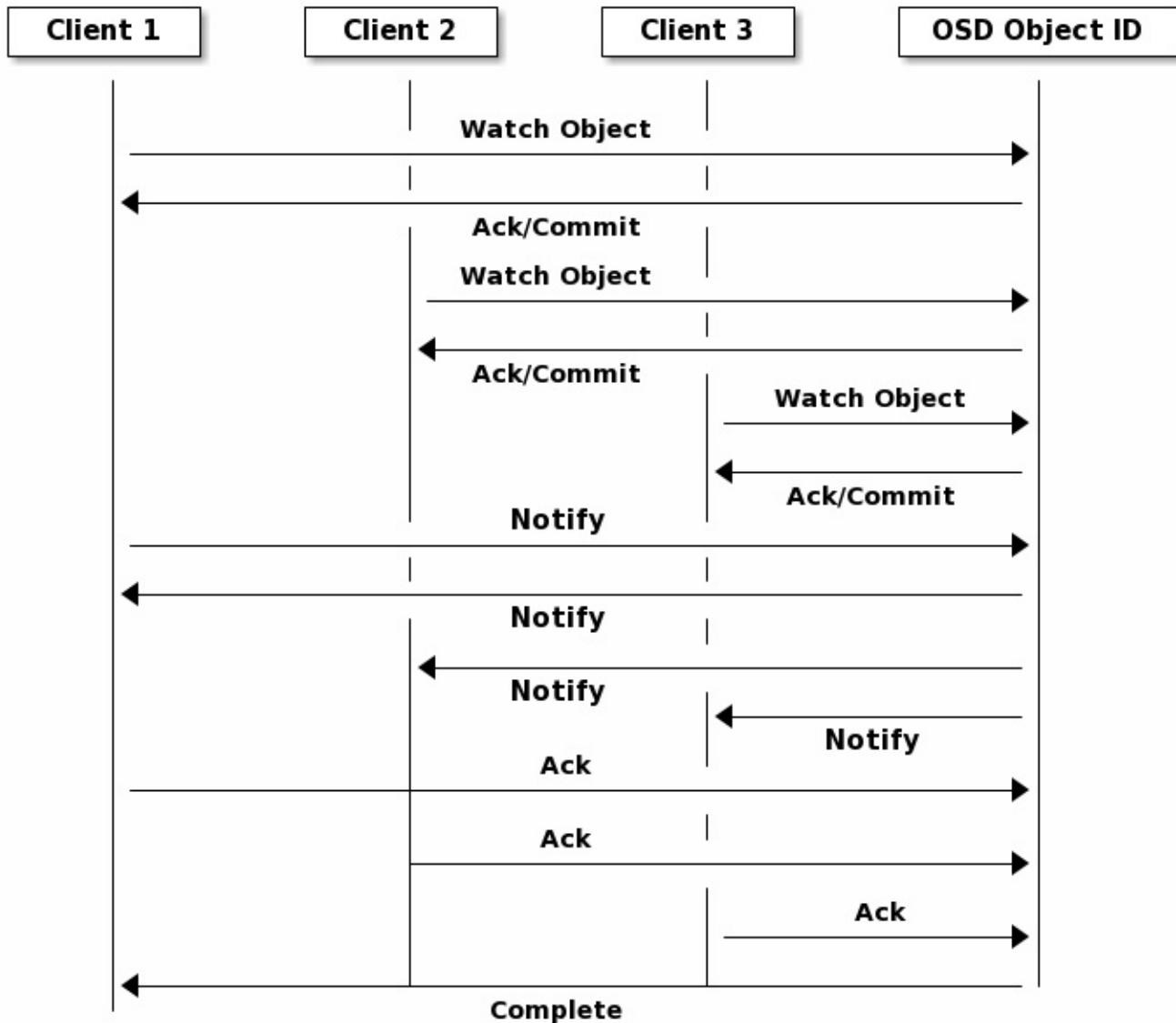
## Native Protocol and `librados`

Modern applications need a simple object storage interface with asynchronous communication capability. The Ceph Storage Cluster provides a simple object storage interface with asynchronous communication capability. The interface provides direct, parallel access to objects throughout the cluster.

- Pool Operations
- Snapshots and Copy-on-write Cloning
- Read/Write Objects - Create or Remove - Entire Object or Byte Range - Append or Truncate
- Create/Set/Get/Remove XATTRs
- Create/Set/Get/Remove Key/Value Pairs
- Compound operations and dual-ack semantics
- Object Classes

## Object Watch/Notify

A client can register a persistent interest with an object and keep a session to the primary OSD open. The client can send a notification message and a payload to all watchers and receive notification when the watchers receive the notification. This enables a client to use any object as a synchronization/communication channel.



## Data Striping

Storage devices have throughput limitations, which impact performance and scalability. So storage systems often support [striping](#)-storing sequential pieces of information across multiple storage devices-to increase throughput and performance. The most common form of data striping comes from [RAID](#). The RAID type most similar to Ceph's striping is [RAID 0](#), or a 'striped volume'. Ceph's striping offers the throughput of RAID 0 striping, the reliability of n-way RAID mirroring and faster recovery.

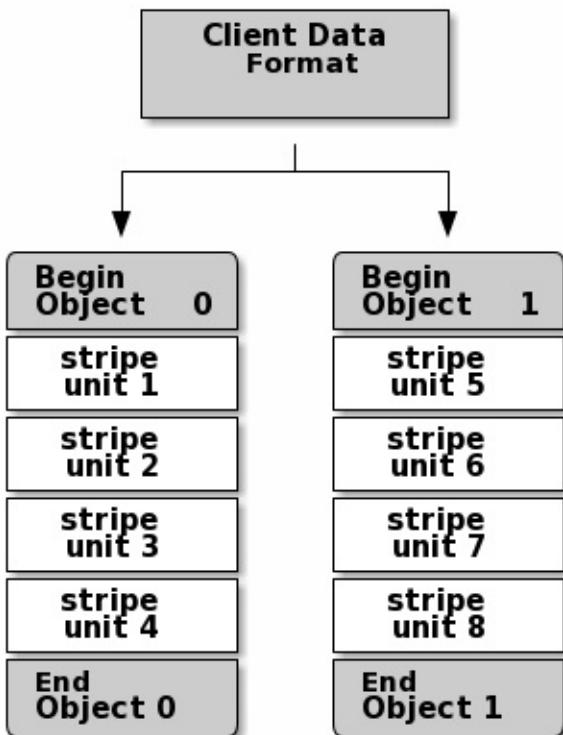
Ceph provides three types of clients: Ceph Block Device, Ceph File System, and Ceph Object Storage. A Ceph Client converts its data from the representation format it provides to its users (a block device image, RESTful objects, CephFS filesystem

directories) into objects for storage in the Ceph Storage Cluster.

### Tip

The objects Ceph stores in the Ceph Storage Cluster are not striped. Ceph Object Storage, Ceph Block Device, and the Ceph File System stripe their data over multiple Ceph Storage Cluster objects. Ceph Clients that write directly to the Ceph Storage Cluster via `librados` must perform the striping (and parallel I/O) for themselves to obtain these benefits.

The simplest Ceph striping format involves a stripe count of 1 object. Ceph Clients write stripe units to a Ceph Storage Cluster object until the object is at its maximum capacity, and then create another object for additional stripes of data. The simplest form of striping may be sufficient for small block device images, S3 or Swift objects and CephFS files. However, this simple form doesn't take maximum advantage of Ceph's ability to distribute data across placement groups, and consequently doesn't improve performance very much. The following diagram depicts the simplest form of striping:



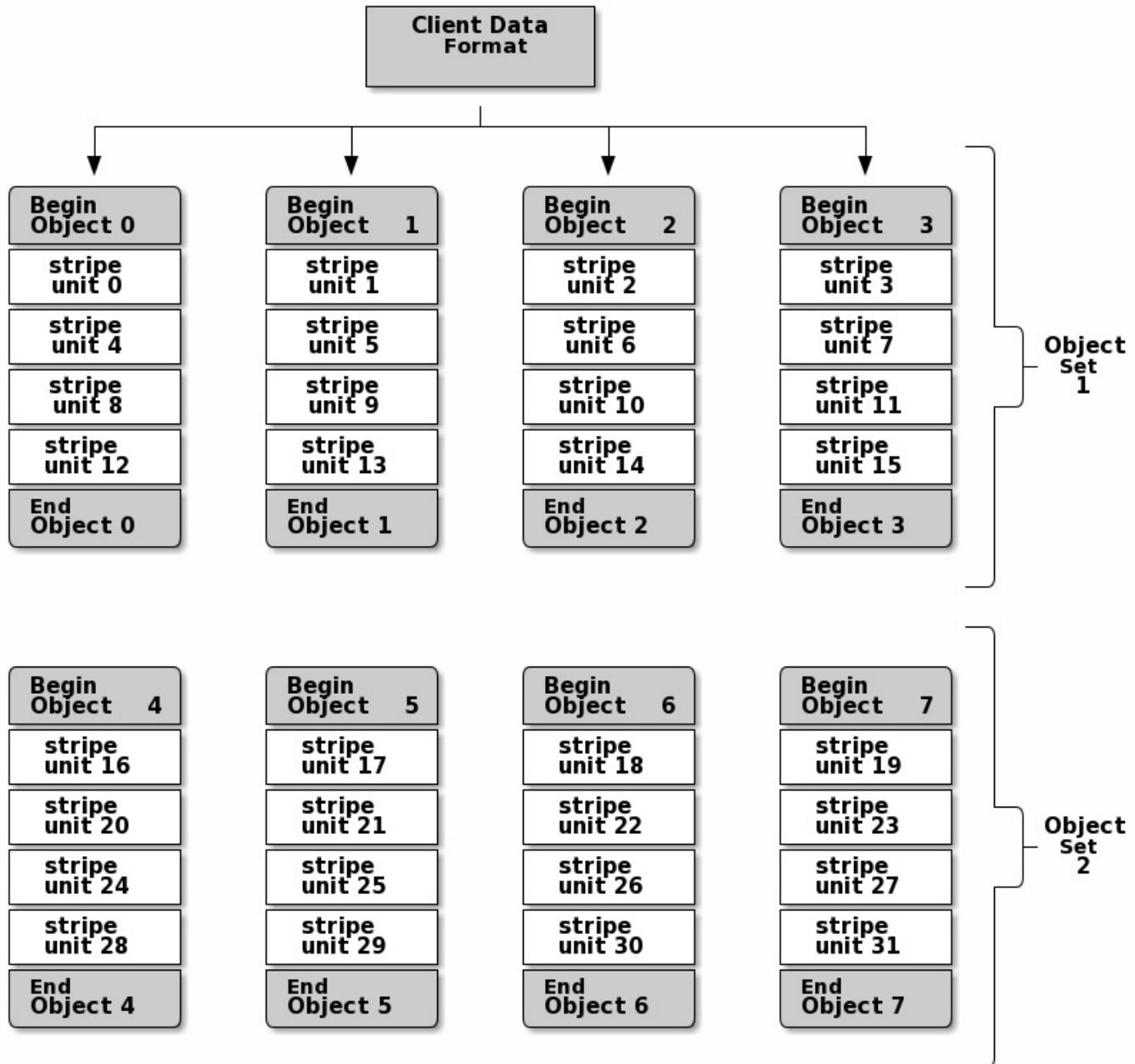
If you anticipate large images sizes, large S3 or Swift objects (e.g., video), or large CephFS directories, you may see considerable read/write performance improvements by striping client data over multiple objects within an object set. Significant write performance occurs when the client writes the stripe units to their corresponding objects in parallel. Since objects get mapped to different placement groups and further mapped to different OSDs, each write occurs in parallel at the maximum write speed. A write to a single drive would be limited by the head movement (e.g. 6ms per seek) and bandwidth of that one device (e.g. 100MB/s). By spreading that write over

multiple objects (which map to different placement groups and OSDs) Ceph can reduce the number of seeks per drive and combine the throughput of multiple drives to achieve much faster write (or read) speeds.

#### Note

Striping is independent of object replicas. Since CRUSH replicates objects across OSDs, stripes get replicated automatically.

In the following diagram, client data gets striped across an object set ( `object set 1` in the following diagram) consisting of 4 objects, where the first stripe unit is `stripe unit 0` in `object 0`, and the fourth stripe unit is `stripe unit 3` in `object 3`. After writing the fourth stripe, the client determines if the object set is full. If the object set is not full, the client begins writing a stripe to the first object again ( `object 0` in the following diagram). If the object set is full, the client creates a new object set ( `object set 2` in the following diagram), and begins writing to the first stripe ( `stripe unit 16` ) in the first object in the new object set ( `object 4` in the diagram below).



Three important variables determine how Ceph stripes data:

- **Object Size:** Objects in the Ceph Storage Cluster have a maximum configurable size (e.g., 2MB, 4MB, etc.). The object size should be large enough to accommodate many stripe units, and should be a multiple of the stripe unit.
- **Stripe Width:** Stripes have a configurable unit size (e.g., 64kb). The Ceph Client divides the data it will write to objects into equally sized stripe units, except for the last stripe unit. A stripe width, should be a fraction of the Object Size so that an object may contain many stripe units.
- **Stripe Count:** The Ceph Client writes a sequence of stripe units over a series of objects determined by the stripe count. The series of objects is called an object set. After the Ceph Client writes to the last object in the object set, it returns to the first object in the object set.

## Important

Test the performance of your striping configuration before putting your cluster into production. You CANNOT change these striping parameters after you stripe the data and write it to objects.

Once the Ceph Client has striped data to stripe units and mapped the stripe units to objects, Ceph's CRUSH algorithm maps the objects to placement groups, and the placement groups to Ceph OSD Daemons before the objects are stored as files on a storage drive.

### Note

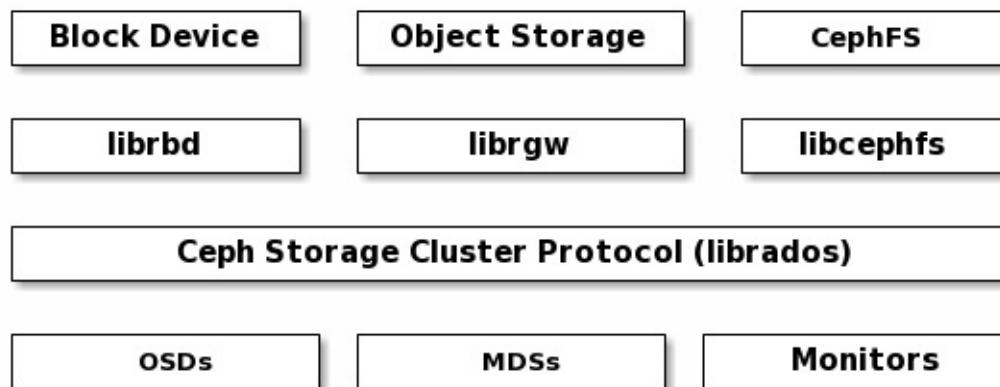
Since a client writes to a single pool, all data striped into objects get mapped to placement groups in the same pool. So they use the same CRUSH map and the same access controls.

## Ceph Clients

Ceph Clients include a number of service interfaces. These include:

- **Block Devices:** The [Ceph Block Device](#) (a.k.a., RBD) service provides resizable, thin-provisioned block devices with snapshotting and cloning. Ceph stripes a block device across the cluster for high performance. Ceph supports both kernel objects (K0) and a QEMU hypervisor that uses `librbd` directly-avoiding the kernel object overhead for virtualized systems.
- **Object Storage:** The [Ceph Object Storage](#) (a.k.a., RGW) service provides RESTful APIs with interfaces that are compatible with Amazon S3 and OpenStack Swift.
- **Filesystem:** The [Ceph File System](#) (CephFS) service provides a POSIX compliant filesystem usable with `mount` or as a filesystem in user space (FUSE).

Ceph can run additional instances of OSDs, MDSS, and monitors for scalability and high availability. The following diagram depicts the high-level architecture.



# Ceph Object Storage

The Ceph Object Storage daemon, `radosgw`, is a FastCGI service that provides a RESTful HTTP API to store objects and metadata. It layers on top of the Ceph Storage Cluster with its own data formats, and maintains its own user database, authentication, and access control. The RADOS Gateway uses a unified namespace, which means you can use either the OpenStack Swift-compatible API or the Amazon S3-compatible API. For example, you can write data using the S3-compatible API with one application and then read data using the Swift-compatible API with another application.

## S3/Swift Objects and Store Cluster Objects Compared

Ceph's Object Storage uses the term *object* to describe the data it stores. S3 and Swift objects are not the same as the objects that Ceph writes to the Ceph Storage Cluster. Ceph Object Storage objects are mapped to Ceph Storage Cluster objects. The S3 and Swift objects do not necessarily correspond in a 1:1 manner with an object stored in the storage cluster. It is possible for an S3 or Swift object to map to multiple Ceph objects.

See [Ceph Object Storage](#) for details.

# Ceph Block Device

A Ceph Block Device stripes a block device image over multiple objects in the Ceph Storage Cluster, where each object gets mapped to a placement group and distributed, and the placement groups are spread across separate `ceph-osd` daemons throughout the cluster.

## Important

Striping allows RBD block devices to perform better than a single server could!

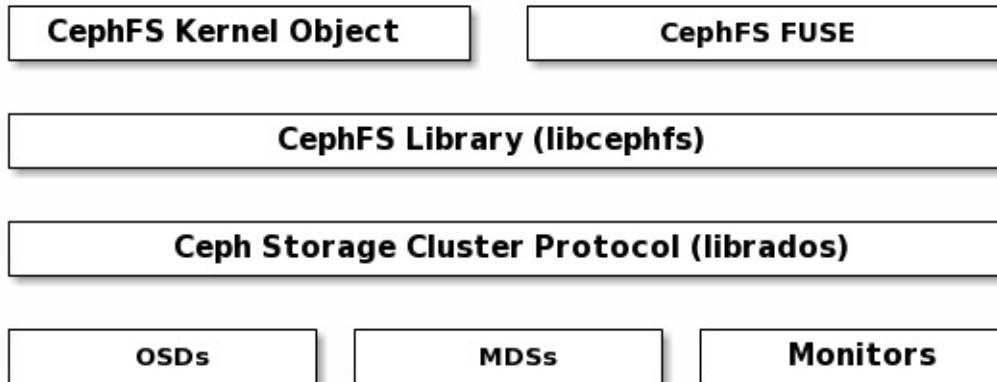
Thin-provisioned snapshottable Ceph Block Devices are an attractive option for virtualization and cloud computing. In virtual machine scenarios, people typically deploy a Ceph Block Device with the `rbd` network storage driver in QEMU/KVM, where the host machine uses `librbd` to provide a block device service to the guest. Many cloud computing stacks use `libvirt` to integrate with hypervisors. You can use thin-provisioned Ceph Block Devices with QEMU and `libvirt` to support OpenStack and CloudStack among other solutions.

While we do not provide `librbd` support with other hypervisors at this time, you may also use Ceph Block Device kernel objects to provide a block device to a client. Other virtualization technologies such as Xen can access the Ceph Block Device kernel object(s). This is done with the command-line tool `rbd`.

# Ceph File System

The Ceph File System (CephFS) provides a POSIX-compliant filesystem as a service that

is layered on top of the object-based Ceph Storage Cluster. CephFS files get mapped to objects that Ceph stores in the Ceph Storage Cluster. Ceph Clients mount a CephFS filesystem as a kernel object or as a Filesystem in User Space (FUSE).



The Ceph File System service includes the Ceph Metadata Server (MDS) deployed with the Ceph Storage cluster. The purpose of the MDS is to store all the filesystem metadata (directories, file ownership, access modes, etc) in high-availability Ceph Metadata Servers where the metadata resides in memory. The reason for the MDS (a daemon called `ceph-mds`) is that simple filesystem operations like listing a directory or changing a directory (`ls`, `cd`) would tax the Ceph OSD Daemons unnecessarily. So separating the metadata from the data means that the Ceph File System can provide high performance services without taxing the Ceph Storage Cluster.

CephFS separates the metadata from the data, storing the metadata in the MDS, and storing the file data in one or more objects in the Ceph Storage Cluster. The Ceph filesystem aims for POSIX compatibility. `ceph-mds` can run as a single process, or it can be distributed out to multiple physical machines, either for high availability or for scalability.

- **High Availability:** The extra `ceph-mds` instances can be standby, ready to take over the duties of any failed `ceph-mds` that was active. This is easy because all the data, including the journal, is stored on RADOS. The transition is triggered automatically by `ceph-mon`.
- **Scalability:** Multiple `ceph-mds` instances can be active, and they will split the directory tree into subtrees (and shards of a single busy directory), effectively balancing the load amongst all active servers.

Combinations of standby and active etc are possible, for example running 3 active `ceph-mds` instances for scaling, and one standby instance for high availability.

# Contributing to Ceph: A Guide for Developers

---

Author

Loic Dachary

Author

Nathan Cutler

License

Creative Commons Attribution Share Alike 3.0 (CC-BY-SA-3.0)

Note

You may also be interested in the [Ceph Internals](#) documentation.

- [Introduction](#)
- [Essentials](#)
- [What is Merged and When](#)
- [Issue tracker](#)
- [Basic workflow](#)
- [Tests: Unit Tests](#)
- [Tests: Integration Tests](#)
- [Running Tests Locally](#)
- [Running Integration Tests using Teuthology](#)
- [Running Tests in the Cloud](#)
- [Ceph Dashboard Developer Documentation \(formerly HACKING.rst\)](#)

# Introduction

---

This guide has two aims. First, it should lower the barrier to entry for software developers who wish to get involved in the Ceph project. Second, it should serve as a reference for Ceph developers.

We assume that readers are already familiar with Ceph (the distributed object store and file system designed to provide excellent performance, reliability and scalability). If not, please refer to the [project website](#) and especially the [publications list](#). Another way to learn about what's happening in Ceph is to check out our [youtube channel](#), where we post Tech Talks, Code walk-throughs and Ceph Developer Monthly recordings.

Since this document is to be consumed by developers, who are assumed to have Internet access, topics covered elsewhere, either within the Ceph documentation or elsewhere on the web, are treated by linking. If you notice that a link is broken or if you know of a better link, please [report it as a bug](#).

# Essentials (tl;dr)

This chapter presents essential information that every Ceph developer needs to know.

## Leads

The Ceph project is led by Sage Weil. In addition, each major project component has its own lead. The following table shows all the leads and their nicks on [GitHub](#):

Scope	Lead	GitHub nick
Ceph	Sage Weil	liewegas
RADOS	Neha Ojha	neha-ojha
RGW	Yehuda Sadeh	yehudasa
RGW	Matt Benjamin	mattbenjamin
RBD	Jason Dillaman	dillaman
CephFS	Patrick Donnelly	batrick
Dashboard	Lenz Grimmer	LenzGr
MON	Joao Luis	jecluis
Build/Ops	Ken Dreyer	ktdreyer
Docs	Zac Dover	zdover23

The Ceph-specific acronyms in the table are explained in [Architecture](#).

## History

See the [History chapter of the Wikipedia article](#).

## Licensing

Ceph is free software.

Unless stated otherwise, the Ceph source code is distributed under the terms of the [LGPL2.1](#) or [LGPL3.0](#). For full details, see the file [COPYING](#) in the top-level directory

of the source-code tree.

## Source code repositories

---

The source code of Ceph lives on [GitHub](#) in a number of repositories below the Ceph “organization”.

A working knowledge of `git` is essential to make a meaningful contribution to the project as a developer.

Although the Ceph “organization” includes several software repositories, this document covers only one: <https://github.com/ceph/ceph>.

## Redmine issue tracker

---

Although [GitHub](#) is used for code, Ceph-related issues (Bugs, Features, Backports, Documentation, etc.) are tracked at <http://tracker.ceph.com>, which is powered by [Redmine](#).

The tracker has a Ceph project with a number of subprojects loosely corresponding to the various architectural components (see [Architecture](#)).

Mere [registration](#) in the tracker automatically grants permissions sufficient to open new issues and comment on existing ones.

To report a bug or propose a new feature, [jump to the Ceph project](#) and click on [New issue](#).

## Mailing lists

---

### Ceph Development Mailing List

The `dev@ceph.io` list is for discussion about the development of Ceph, its interoperability with other technology, and the operations of the project itself.

The email discussion list for Ceph development is open to all. Subscribe by sending a message to `dev-request@ceph.io` with the following line in the body of the message:

```
1. subscribe ceph-devel
```

### Ceph Client Patch Review Mailing List

The `ceph-devel@vger.kernel.org` list is for discussion and patch review for the Linux kernel Ceph client component. Note that this list used to be an all-encompassing list for developers. When searching the archives, remember that this list contains the generic

devel-ceph archives before mid-2018.

Subscribe to the list covering the Linux kernel Ceph client component by sending a message to [majordomo@vger.kernel.org](mailto:majordomo@vger.kernel.org) with the following line in the body of the message:

```
1. subscribe ceph-devel
```

## Other Ceph Mailing Lists

There are also [other Ceph-related mailing lists](#).

## IRC

In addition to mailing lists, the Ceph community also communicates in real time using [Internet Relay Chat](#).

See <https://ceph.com/irc/> for how to set up your IRC client and a list of channels.

## Submitting patches

The canonical instructions for submitting patches are contained in the file [CONTRIBUTING.rst](#) in the top-level directory of the source-code tree. There may be some overlap between this guide and that file.

All newcomers are encouraged to read that file carefully.

## Building from source

See instructions at [Build Ceph](#).

## Using ccache to speed up local builds

[ccache](#) can make the process of rebuilding the ceph source tree faster.

Before you use [ccache](#) to speed up your rebuilds of the ceph source tree, make sure that your source tree is clean and will produce no build failures. When you have a clean source tree, you can confidently use [ccache](#), secure in the knowledge that you're not using a dirty tree.

Old build artifacts can cause build failures. You might introduce these artifacts unknowingly when switching from one branch to another. If you see build errors when you attempt a local build, follow the procedure below to clean your source tree.

## Cleaning the Source Tree

```
1. make clean
```

## Note

The following commands will remove everything in the source tree that isn't tracked by git. Make sure to back up your log files and configuration options before running these commands.

```
1. git clean -fdx; git submodule foreach git clean -fdx
```

## Building Ceph with ccache

`ccache` is available as a package in most distros. To build ceph with ccache, run the following command.

```
1. cmake -DWITH_CCACHE=ON ..
```

## Using ccache to Speed Up Build Times

`ccache` can be used for speeding up all builds of the system. For more details, refer to the [run modes](#) section of the ccache manual. The default settings of `ccache` can be displayed with the `ccache -s` command.

## Note

We recommend overriding the `max_size`. The default is 10G. Use a larger value, like 25G. Refer to the [configuration](#) section of the ccache manual for more information.

To further increase the cache hit rate and reduce compile times in a development environment, set the version information and build timestamps to fixed values. This makes it unnecessary to rebuild the binaries that contain this information.

This can be achieved by adding the following settings to the `ccache` configuration file `ccache.conf` :

```
1. sloppiness = time_macros
2. run_second_cpp = true
```

Now, set the environment variable `SOURCE_DATE_EPOCH` to a fixed value (a UNIX timestamp) and set `ENABLE_GIT_VERSION` to `OFF` when running `cmake` :

```
1. export SOURCE_DATE_EPOCH=946684800
2. cmake -DWITH_CCACHE=ON -DENABLE_GIT_VERSION=OFF ..
```

## Note

Binaries produced with these build options are not suitable for production or debugging purposes, as they do not contain the correct build time and git version information.

## Development-mode cluster

---

See [Developer Guide \(Quick\)](#).

## Kubernetes/Rook development cluster

---

See [Hacking on Ceph in Kubernetes with Rook](#)

## Backporting

---

All bugfixes should be merged to the `master` branch before being backported. To flag a bugfix for backporting, make sure it has a [tracker issue](#) associated with it and set the `Backport` field to a comma-separated list of previous releases (e.g. "hammer,jewel") that you think need the backport. The rest (including the actual backporting) will be taken care of by the [Stable Releases and Backports](#) team.

## Guidance for use of cluster log

---

If your patches emit messages to the Ceph cluster log, please consult this: [Use of the cluster log](#).

# Commit merging: scope and cadence

---

Commits are merged into branches according to criteria specific to each phase of the Ceph release lifecycle. This chapter codifies these criteria.

## Development releases (i.e. x.0.z)

---

### What ?

- Features
- Bug fixes

### Where ?

Features are merged to the *master* branch. Bug fixes should be merged to the corresponding named branch (e.g. *nautilus* for 14.0.z, *pacific* for 16.0.z, etc.). However, this is not mandatory - bug fixes and documentation enhancements can be merged to the *master* branch as well, since the *master* branch is itself occasionally merged to the named branch during the development releases phase. In either case, if a bug fix is important it can also be flagged for backport to one or more previous stable releases.

### When ?

After each stable release, candidate branches for previous releases enter phase 2 (see below). For example: the *jewel* named branch was created when the *infernalis* release candidates entered phase 2. From this point on, *master* was no longer associated with *infernalis*. After the named branch of the next stable release is created, *master* will be occasionally merged into it.

## Branch merges

- The latest stable release branch is merged periodically into *master*.
- The *master* branch is merged periodically into the branch of the stable release.
- The *master* is merged into the stable release branch immediately after each development (x.0.z) release.

## Stable release candidates (i.e. x.1.z) phase 1

---

### What ?

- Bug fixes only

## Where ?

The stable release branch (e.g. *jewel* for 10.0.z, *luminous* for 12.0.z, etc.) or *master*. Bug fixes should be merged to the named branch corresponding to the stable release candidate (e.g. *jewel* for 10.1.z) or to *master*. During this phase, all commits to *master* will be merged to the named branch, and vice versa. In other words, it makes no difference whether a commit is merged to the named branch or to *master* - it will make it into the next release candidate either way.

## When ?

After the first stable release candidate is published, i.e. after the x.1.0 tag is set in the release branch.

## Branch merges

- The stable release branch is merged periodically into *master*.
- The *master* branch is merged periodically into the stable release branch.
- The *master* branch is merged into the stable release branch immediately after each x.1.z release candidate.

## Stable release candidates (i.e. x.1.z) phase 2

---

## What ?

- Bug fixes only

## Where ?

The stable release branch (e.g. *mimic* for 13.0.z, *octopus* for 15.0.z ,etc.). During this phase, all commits to the named branch will be merged into *master*. Cherry-picking to the named branch during release candidate phase 2 is performed manually since the official backporting process begins only when the release is pronounced "stable".

## When ?

After Sage Weil announces that it is time for phase 2 to happen.

## Branch merges

- The stable release branch is occasionally merged into *master*.

## Stable releases (i.e. x.2.z)

---

### What ?

- Bug fixes
- Features are ~~sometime~~ accepted
- Commits should be cherry-picked from *master* when possible
- Commits that are not cherry-picked from *master* must pertain to a bug unique to the stable release
- See also the [backport HOWTO](#) document

### Where ?

The stable release branch (*hammer* for 0.94.x, *infernalis* for 9.2.x, etc.)

### When ?

After the stable release is published, i.e. after the “vx.2.0” tag is set in the release branch.

## Branch merges

Never

# Issue Tracker

See [Redmine Issue Tracker](#) for a brief introduction to the Ceph Issue Tracker.

Ceph developers use the issue tracker to

1. keep track of issues - bugs, fix requests, feature requests, backport requests, etc.
2. communicate with other developers and keep them informed as work on the issues progresses.

## Issue tracker conventions

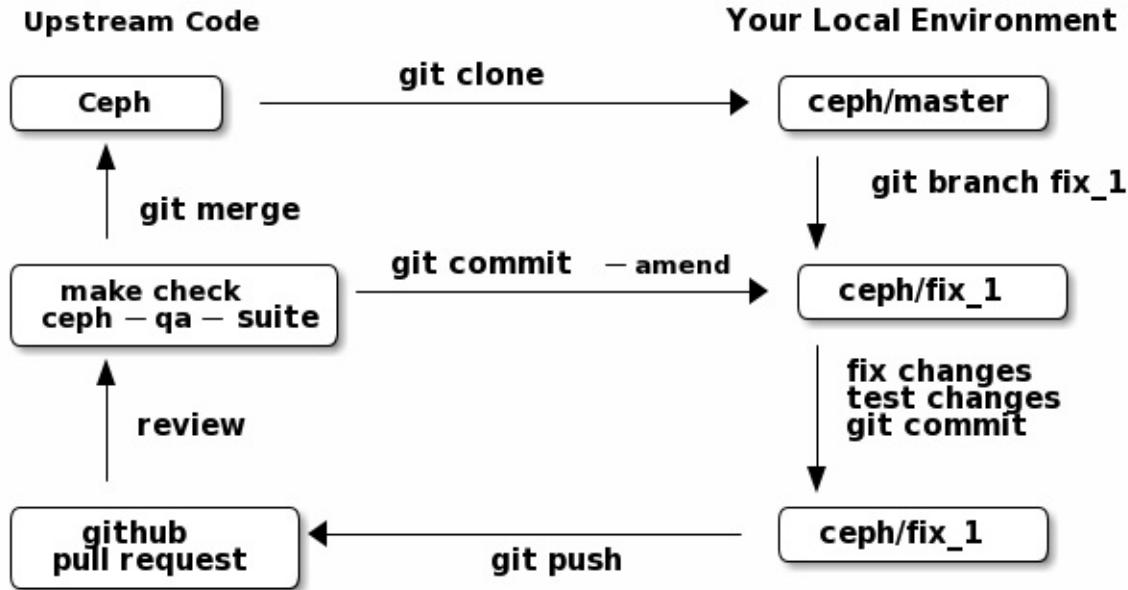
When you start working on an existing issue, it's nice to let the other developers know this - to avoid duplication of labor. Typically, this is done by changing the `Assignee` field (to yourself) and changing the `Status` to *In progress*. Newcomers to the Ceph community typically do not have sufficient privileges to update these fields, however: they can simply update the issue with a brief note.

Meanings of some commonly used statuses

Status	Meaning
New	Initial status
In Progress	Somebody is working on it
Need Review	Pull request is open with a fix
Pending Backport	Fix has been merged, backport(s) pending
Resolved	Fix and backports (if any) have been merged

# Basic Workflow

The following chart illustrates the basic Ceph development workflow:



The below explanation is written with the assumption that you, the reader, are a new contributor who has an idea for a bugfix or enhancement, but do not know exactly how to proceed. Watch the [Getting Started with Ceph Development](#) video for a practical summary of this workflow.

## Update the tracker

Before you start, you should know the [Issue Tracker](#) (Redmine) number of the bug you intend to fix. If there is no tracker issue, now is the time to create one for code changes. Straightforward documentation cleanup does not necessarily require a corresponding tracker issue. However, an issue (ticket) should be created if one is adding new documentation chapters or files, or for other substantial changes.

The tracker ticket serves to explain the issue (bug) to your fellow Ceph developers and keep them informed as you make progress toward resolution. To this end, please provide a descriptive title and write appropriate information and details into the description. When composing the ticket's title, consider “If I want to search for this ticket two years from now, what keywords will I search for?”

If you have sufficient tracker permissions, assign the bug to yourself by setting the `Assignee` field. If your tracker permissions have not been elevated, simply add a comment with a short message like “I am working on this issue”.

# Upstream code

This section, and the ones that follow, correspond to nodes in the above chart.

The upstream code is found at <https://github.com/ceph/ceph.git>, which is known as the “upstream repo”, or simply “upstream”. As the chart shows, we will make a local copy of this repository, modify it, test our modifications, then submit the modifications for review and merging.

A local copy of the upstream code is made by

1. Forking the upstream repo on GitHub, and
2. Cloning your fork to make a local working copy

See the [GitHub documentation](#) for detailed instructions on forking. In short, if your GitHub username is “mygithubaccount”, your fork of the upstream repo will appear at <https://github.com/mygithubaccount/ceph>. Once you have created your fork, clone it by running:

```
1. git clone https://github.com/mygithubaccount/ceph
```

While it is possible to clone the upstream repo directly, for the Ceph workflow you must fork it first. Forking is what enables us to open a [GitHub pull request](#).

For more information on using GitHub, refer to [GitHub Help](#).

## Local environment

In the local environment created in the previous step, you now have a copy of the `master` branch in `remotes/origin/master`. This fork (<https://github.com/mygithubaccount/ceph.git>) is frozen in time and the upstream repo (<https://github.com/ceph/ceph.git>, typically abbreviated to `ceph/ceph.git`) is updated frequently by other contributors, you must sync your fork periodically. Failure to do so may result in your commits and pull requests failing to merge because they refer to file contents that have since changed.

First, ensure that you have properly configured your local git environment with your name and email address. Skip this step if you have already configured this information.

```
1. git config user.name "FIRST_NAME LAST_NAME"  
2. git config user.email "MY_NAME@example.com"
```

Now add the upstream repo as a “remote” and fetch it:

```
1. git remote add ceph https://github.com/ceph/ceph.git
```

```
2. git fetch ceph
```

Fetching downloads all objects (commits, branches) that were added since the last sync. After running these commands, all the branches from `ceph/ceph.git` are downloaded to the local git repo as `remotes/ceph/$BRANCH_NAME` and can be referenced as `ceph/$BRANCH_NAME` in local git commands.

For example, your local `master` branch can be reset to the upstream Ceph `master` branch by running

```
1. git fetch ceph
2. git checkout master
3. git reset --hard ceph/master
```

Finally, the `master` branch of your fork is synced to the upstream master by

```
1. git push -u origin master
```

## Bugfix branch

Next, create a branch for your bugfix:

```
1. git checkout master
2. git checkout -b fix_1
3. git push -u origin fix_1
```

This creates a `fix_1` branch locally and in our GitHub fork. At this point, the `fix_1` branch is identical to the `master` branch, but not for long! You are now ready to modify the code. Be careful to always run `git checkout master` first, otherwise you may find commits from an unrelated branch mixed with your new work.

## Fix bug locally

Now change the status of the tracker issue to “In progress” to communicate to other Ceph contributors that you have begun working on a fix. This helps avoid duplication of effort. If you don’t have permission to change that field, your previous comment that you are working on the issue is sufficient.

Your fix may be very simple and require only minimal testing. More likely, this will be an iterative process involving trial and error, not to mention skill. An explanation of how to fix bugs is beyond the scope of this document. Instead, we focus on the mechanics of the process in the context of the Ceph project.

For a detailed discussion of the tools available for validating bugfixes, see the chapters on testing.

For now, let us just assume that you have finished work on the bugfix, that you have tested, and that you believe it works. Commit the changes to your local branch using the `--signoff` option

```
1. git commit -as
```

and push the changes to your fork

```
1. git push origin fix_1
```

## GitHub pull request

The next step is to open a GitHub pull request (PR). This makes your bugfix visible to the community of Ceph contributors. They will review it and may perform additional testing and / or request changes.

This is the point where you “go public” with your modifications. Be prepared to receive suggestions and constructive criticism in the form of comments within the PR. Don’t worry! The Ceph project is a friendly place!

If you are uncertain how to create and manage pull requests, you may read [this GitHub pull request tutorial](#).

For ideas on what constitutes a “good” pull request, see the [Git Commit Good Practice](#) article at the [OpenStack Project Wiki](#).

and our own [Submitting Patches](#) document.

Once your pull request (PR) is opened, update the [Issue Tracker](#) by adding a comment directing other contributors to your PR. The comment can be as simple as:

```
1. *PR*: https://github.com/ceph/ceph/pull/$NUMBER_OF_YOUR_PULL_REQUEST
```

## Automated PR validation

When your PR is created or updated, the Ceph project’s [Continuous Integration \(CI\)](#) infrastructure will test it automatically. At the time of this writing (September 2020), the automated CI testing included five tests to check that the commits in the PR are properly signed (see [Submitting patches](#)), to check that the documentation builds, to check that the submodules are unmodified, to check that the API is in order, and a [What does “make check” mean?](#) test. Additional tests may be performed depending on which files are modified by your PR.

The [What does “make check” mean?](#), builds the PR and runs it through a battery of tests. These tests run on servers operated by the Ceph Continuous Integration (CI) team. When the tests complete, the result will be shown on GitHub in the pull request

itself.

You can (and should) also test your modifications before you open a PR. Refer to the chapters on testing for details.

## Notes on PR make check test

The GitHub [make check](#) test is driven by a Jenkins instance.

Jenkins merges your PR branch into the latest version of the base branch before starting tests. This means that you don't have to rebase the PR to pick up any fixes.

You can trigger PR tests at any time by adding a comment to the PR - the comment should contain the string "test this please". Since a human subscribed to the PR might interpret that as a request for him or her to test the PR, we recommend that you address Jenkins directly. For example, write "jenkins retest this please". For efficiency a single re-test can also be requested with e.g. "jenkins test signed". For reference, a list of these requests is automatically added to the end of each new PR's description.

If there is a build failure and you aren't sure what caused it, check the [make check](#) log. To access it, click on the "details" (next to the [make check](#) test in the PR) link to enter the Jenkins web GUI. Then click on "Console Output" (on the left).

Jenkins is configured to search logs for strings known to have been associated with [make check](#) failures in the past. However, there is no guarantee that these known strings are associated with any given [make check](#) failure. You'll have to read through the log to determine the cause of your specific failure.

## Integration tests AKA ceph-qa-suite

---

Since Ceph is complex, it may be necessary to test your fix to see how it behaves on real clusters running on physical or virtual hardware. Tests designed for this purpose live in the [ceph/qa sub-directory](#) and are run via the [teuthology framework](#).

The Ceph community has access to the [Sepia lab](#) where [Testing - Integration Tests](#) can be run on physical hardware. Other developers may add tags like "needs-qa" to your PR. This allows PRs that need testing to be merged into a single branch and tested all at the same time. Since teuthology suites can take hours (even days in some cases) to run, this can save a lot of time.

To request access to the Sepia lab, start [here](#).

Integration testing is discussed in more detail in the [Testing - Integration Tests](#) chapter.

## Code review

Once your bugfix has been thoroughly tested, or even during this process, it will be subjected to code review by other developers. This typically takes the form of comments in the PR itself, but can be supplemented by discussions on [IRC](#) and the [Mailing lists](#).

## Amending your PR

While your PR is going through testing and [Code Review](#), you can modify it at any time by editing files in your local branch.

After updates are committed locally (to the `fix_1` branch in our example), they need to be pushed to GitHub so they appear in the PR.

Modifying the PR is done by adding commits to the `fix_1` branch upon which it is based, often followed by rebasing to modify the branch's git history. See [this tutorial](#) for a good introduction to rebasing. When you are done with your modifications, you will need to force push your branch with:

```
1. git push --force origin fix_1
```

Why do we take these extra steps instead of simply adding additional commits to the PR? It is best practice for a PR to consist of a single commit; this makes for clean history, eases peer review of your changes, and facilitates merges. In rare circumstances it also makes it easier to cleanly revert changes.

## Merge

The bugfix process completes when a project lead merges your PR.

When this happens, it is a signal for you (or the lead who merged the PR) to change the [Issue Tracker](#) status to "Resolved". Some issues may be flagged for backporting, in which case the status should be changed to "Pending Backport" (see the [Backporting](#) chapter for details).

See also [Commit merging: scope and cadence](#) for more information on merging.

## Proper Merge Commit Format

This is the most basic form of a merge commit:

```
1. doc/component: title of the commit  
2.  
3. Reviewed-by: Reviewer Name <rname@example.com>
```

This consists of two parts:

1. The title of the commit / PR to be merged.
2. The name and email address of the reviewer. Enclose the reviewer's email address in angle brackets.

## Using .githubmap to Find a Reviewer's Email Address

If you cannot find the email address of the reviewer on his or her GitHub page, you can look it up in the `.githubmap` file, which can be found in the repository at `/ceph/.githubmap`.

## Using "git log" to find a Reviewer's Email Address

If you cannot find a reviewer's email address by using the above methods, you can search the git log for their email address. Reviewers are likely to have committed something before. If they have made previous contributions, the git log will probably contain their email address.

Use the following command

```
1. git log
```

## Using ptl-tool to Generate Merge Commits

Another method of generating merge commits involves using Patrick Donnelly's `ptl-tool` pull commits. This tool can be found at `/ceph/src/script/ptl-tool.py`. Merge commits that have been generated by the `ptl-tool` have the following form:

```
1. Merge PR #36257 into master
2. * refs/pull/36257/head:
3.     client: move client_lock to _unmount()
4.     client: add timer_lock support
5. Reviewed-by: Patrick Donnelly <pdonnell@redhat.com>
```

# Testing - unit tests

The Ceph GitHub repository has two types of tests: unit tests (also called `make check` tests) and integration tests. Strictly speaking, the `make check` tests are not “unit tests”, but rather tests that can be run easily on a single build machine after compiling Ceph from source, whereas integration tests require package installation and multi-machine clusters to run.

## What does “make check” mean?

After compiling Ceph, the code can be run through a battery of tests. For historical reasons, this is often referred to as `make check` even though the actual command used to run the tests is now `ctest`. For inclusion in this group of tests, a test must:

- bind ports that do not conflict with other tests
- not require root access
- not require more than one machine to run
- complete within a few minutes

For the sake of simplicity, this class of tests is referred to as “make check tests” or “unit tests”. This is meant to distinguish these tests from the more complex “integration tests” that are run via the [teuthology framework](#).

While it is possible to run `ctest` directly, it can be tricky to correctly set up your environment. Fortunately, a script is provided to make it easier run the unit tests on your code. It can be run from the top-level directory of the Ceph source tree by invoking:

```
1. ... prompt:: bash $
```

You will need a minimum of 8GB of RAM and 32GB of free drive space for this command to complete successfully on x86\_64; other architectures may have different requirements. Depending on your hardware, it can take from twenty minutes to three hours to complete, but it’s worth the wait.

## How unit tests are declared

Unit tests are declared in the `CMakeLists.txt` file, which is found in the `./src` directory. The `add_ceph_test` and `add_ceph_unittest` CMake functions are used to declare unit tests. `add_ceph_test` and `add_ceph_unittest` are themselves defined in `./cmake/modules/AddCephTest.cmake`.

Some unit tests are scripts and other unit tests are binaries that are compiled during the build process.

- `add_ceph_test` function - used to declare unit test scripts
- `add_ceph_unittest` function - used for unit test binaries

## Unit testing of CLI tools

Some of the CLI tools are tested using special files ending with the extension `.t` and stored under `./src/test/cli`. These tests are run using a tool called `cram` via a shell script `./src/test/run-cli-tests`. `cram` tests that are not suitable for `make check` may also be run by teuthology using the `cram task`.

## Tox based testing of python modules

Most python modules can be found under `./src/pybind/`.

Many modules use `tox` to run their unit tests. `tox` itself is a generic virtualenv management and test command line tool.

To find out quickly if `tox` can be run you can either just try to run `tox` or check for the existence of a `tox.ini` file.

Currently the following modules use `tox`:

- Cephadm (`./src/pybind/mgr/cephadm`)
- Insights (`./src/pybind/mgr/insights`)
- Manager core (`./src/pybind/mgr`)
- Dashboard (`./src/pybind/mgr/dashboard`)
- Python common (`./src/python-common/tox.ini`)

Most `tox` configurations support multiple environments and tasks. You can see which are supported by examining the `envlist` assignment within `tox.ini`. To run `tox`, just execute `tox` in the directory where `tox.ini` is found. If no environments are specified with e.g. `-e $env1,$env2`, all environments will be run. Jenkins will run `tox` by executing `run_tox.sh` which is under `./src/script`.

Here some examples from the Ceph Dashboard on how to specify environments and run options:

1. ## Run Python 2+3 tests+lint commands:
2. \$ tox -e py27,py3,lint,check
- 3.
4. ## Run Python 3 tests+lint commands:
5. \$ tox -e py3,lint,check

```
6.  
7. ## To run it like Jenkins would do  
8. $ ../../script/run_tox.sh --tox-env py27,py3 lint check  
9. $ ../../script/run_tox.sh --tox-env py3 lint check
```

## Manager core unit tests

Currently only `doctests` inside `mgr_util.py` are run.

To add test additional files inside the core of the manager, add them at the end of the line that includes `mgr_util.py` within `tox.ini`.

## Unit test caveats

1. Unlike the various Ceph daemons and `ceph-fuse`, unit tests are linked against the default memory allocator (glibc) unless explicitly linked against something else. This enables tools like `valgrind` to be used in the tests.

# Testing - Integration Tests

---

Ceph has two types of tests: `make check` tests and integration tests. When a test requires multiple machines, root access or lasts for a longer time (for example, to simulate a realistic Ceph deployment), it is deemed to be an integration test. Integration tests are organized into “suites”, which are defined in the `ceph/qa` sub-directory and run with the `teuthology-suite` command.

The `teuthology-suite` command is part of the `teuthology framework`. In the sections that follow we attempt to provide a detailed introduction to that framework from the perspective of a beginning Ceph developer.

## Teuthology consumes packages

---

It may take some time to understand the significance of this fact, but it is very significant. It means that automated tests can be conducted on multiple platforms using the same packages (RPM, DEB) that can be installed on any machine running those platforms.

Teuthology has a [list of platforms that it supports](#) (as of September 2020 the list consisted of “RHEL/CentOS 8” and “Ubuntu 18.04”). It expects to be provided pre-built Ceph packages for these platforms. Teuthology deploys these platforms on machines (bare-metal or cloud-provisioned), installs the packages on them, and deploys Ceph clusters on them - all as called for by the test.

## The Nightlies

---

A number of integration tests are run on a regular basis in the [Sepia lab](#) against the official Ceph repositories (on the `master` development branch and the stable branches). Traditionally, these tests are called “the nightlies” because the Ceph core developers used to live and work in the same time zone and from their perspective the tests were run overnight.

The results of the nightlies are published at <http://pulpito.ceph.com/>. The developer nick shows in the test results URL and in the first column of the Pulpito dashboard. The results are also reported on the [ceph-qa mailing list](#) for analysis.

## Testing Priority

---

The `teuthology-suite` command includes an almost mandatory option `-p <N>` which specifies the priority of the jobs submitted to the queue. The lower the value of `N`, the higher the priority. The option is almost mandatory because the default is `1000` which matches the priority of the nightlies. Nightlies are often half-finished and cancelled due to the volume of testing done so your jobs may never finish. Therefore,

it is common to select a priority less than 1000.

Job priority should be selected based on the following recommendations:

- **Priority < 10:** Use this if the sky is falling and some group of tests must be run ASAP.
- **10 <= Priority < 50:** Use this if your tests are urgent and blocking other important development.
- **50 <= Priority < 75:** Use this if you are testing a particular feature/fix and running fewer than about 25 jobs. This range can also be used for urgent release testing.
- **75 <= Priority < 100:** Tech Leads will regularly schedule integration tests with this priority to verify pull requests against master.
- **100 <= Priority < 150:** This priority is to be used for QE validation of point releases.
- **150 <= Priority < 200:** Use this priority for 100 jobs or fewer of a particular feature/fix that you'd like results on in a day or so.
- **200 <= Priority < 1000:** Use this priority for large test runs that can be done over the course of a week.

In case you don't know how many jobs would be triggered by `teuthology-suite` command, use `--dry-run` to get a count first and then issue `teuthology-suite` command again, this time without `--dry-run` and with `-p` and an appropriate number as an argument to it.

To skip the priority check, use `--force-priority`. In order to be sensitive to the runs of other developers who also need to do testing, please use it in emergency only.

## Suites Inventory

---

The `suites` directory of the `ceph/qa` sub-directory contains all the integration tests, for all the Ceph components.

### `ceph-deploy`

install a Ceph cluster with `ceph-deploy` ([ceph-deploy man page](#))

### `dummy`

get a machine, do nothing and return success (commonly used to verify the [Testing - Integration Tests](#) infrastructure works as expected)

### `fs`

test CephFS mounted using FUSE

## kcephfs

test CephFS mounted using kernel

## krbd

test the RBD kernel module

## multimds

test CephFS with multiple MDSS

## powercycle

verify the Ceph cluster behaves when machines are powered off and on again

## rados

run Ceph clusters including OSDs and MONs, under various conditions of stress

## rbd

run RBD tests using actual Ceph clusters, with and without qemu

## rgw

run RGW tests using actual Ceph clusters

## smoke

run tests that exercise the Ceph API with an actual Ceph cluster

## teuthology

verify that teuthology can run integration tests, with and without OpenStack

## upgrade

for various versions of Ceph, verify that upgrades can happen without disrupting an ongoing workload

# teuthology-describe-tests

---

In February 2016, a new feature called [teuthology-describe-tests](#) was added to the [teuthology framework](#) to facilitate documentation and better understanding of integration tests ([feature announcement](#)).

The upshot is that tests can be documented by embedding [meta:](#) annotations in the yaml files used to define the tests. The results can be seen in the [ceph-qa-suite wiki](#).

Since this is a new feature, many yaml files have yet to be annotated. Developers are encouraged to improve the documentation, in terms of both coverage and quality.

# How integration tests are run

Given that - as a new Ceph developer - you will typically not have access to the [Sepia lab](#), you may rightly ask how you can run the integration tests in your own environment.

One option is to set up a teuthology cluster on bare metal. Though this is a non-trivial task, it is possible. Here are [some notes](#) to get you started if you decide to go this route.

If you have access to an OpenStack tenant, you have another option: the [teuthology framework](#) has an OpenStack backend, which is documented [here](#). This OpenStack backend can build packages from a given git commit or branch, provision VMs, install the packages and run integration tests on those VMs. This process is controlled using a tool called `ceph-workbench ceph-qa-suite`. This tool also automates publishing of test results at <http://teuthology-logs.public.ceph.com>.

Running integration tests on your code contributions and publishing the results allows reviewers to verify that changes to the code base do not cause regressions, or to analyze test failures when they do occur.

Every teuthology cluster, whether bare-metal or cloud-provisioned, has a so-called “teuthology machine” from which tests suites are triggered using the `teuthology-suite` command.

A detailed and up-to-date description of each `teuthology-suite` option is available by running the following command on the teuthology machine

```
1. teuthology-suite --help
```

# How integration tests are defined

Integration tests are defined by yaml files found in the `suites` subdirectory of the [ceph/qa sub-directory](#) and implemented by python code found in the `tasks` subdirectory. Some tests (“standalone tests”) are defined in a single yaml file, while other tests are defined by a directory tree containing yaml files that are combined, at runtime, into a larger yaml file.

## Reading a standalone test

Let us first examine a standalone test, or “singleton”.

Here is a commented example using the integration test [rados/singleton/all/admin-socket.yaml](#)

```
1. roles:
```

```

2.   - - mon.a
3.   - osd.0
4.   - osd.1
5. tasks:
6. - install:
7. - ceph:
8. - admin_socket:
9.   osd.0:
10.    version:
11.    git_version:
12.    help:
13.    config show:
14.    config set filestore_dump_file /tmp/foo:
15.    perf dump:
16.    perf schema:

```

The `roles` array determines the composition of the cluster (how many MONs, OSDs, etc.) on which this test is designed to run, as well as how these roles will be distributed over the machines in the testing cluster. In this case, there is only one element in the top-level array: therefore, only one machine is allocated to the test. The nested array declares that this machine shall run a MON with id `a` (that is the `mon.a` in the list of roles) and two OSDs (`osd.0` and `osd.1`).

The body of the test is in the `tasks` array: each element is evaluated in order, causing the corresponding python file found in the `tasks` subdirectory of the [teuthology repository](#) or [ceph/qa sub-directory](#) to be run. “Running” in this case means calling the `task()` function defined in that file.

In this case, the `install` task comes first. It installs the Ceph packages on each machine (as defined by the `roles` array). A full description of the `install` task is [found in the python file](#) (search for “def task”).

The `ceph` task, which is documented [here](#) (again, search for “def task”), starts OSDs and MONs (and possibly MDSs as well) as required by the `roles` array. In this example, it will start one MON (`mon.a`) and two OSDs (`osd.0` and `osd.1`), all on the same machine. Control moves to the next task when the Ceph cluster reaches `HEALTH_OK` state.

The next task is `admin_socket` ([source code](#)). The parameter of the `admin_socket` task (and any other task) is a structure which is interpreted as documented in the task. In this example the parameter is a set of commands to be sent to the admin socket of `osd.0`. The task verifies that each of them returns on success (i.e. exit code zero).

This test can be run with

```
1. teuthology-suite --machine-type smithi --suite rados/singleton/all/admin-socket.yaml fs/ext4.yaml
```

## Test descriptions

Each test has a “test description”, which is similar to a directory path, but not the same. In the case of a standalone test, like the one in [Reading a standalone test](#), the test description is identical to the relative path (starting from the `suites/` directory of the `ceph/qa sub-directory`) of the yaml file defining the test.

Much more commonly, tests are defined not by a single yaml file, but by a directory tree of yaml files. At runtime, the tree is walked and all yaml files (facets) are combined into larger yaml “programs” that define the tests. A full listing of the yaml defining the test is included at the beginning of every test log.

In these cases, the description of each test consists of the subdirectory under `suites/` containing the yaml facets, followed by an expression in curly braces (`{ }`) consisting of a list of yaml facets in order of concatenation. For instance the test description:

```
1. ceph-deploy/basic/{distros/centos_7.0.yaml tasks/ceph-deploy.yaml}
```

signifies the concatenation of two files:

- `ceph-deploy/basic/distros/centos_7.0.yaml`
- `ceph-deploy/basic/tasks/ceph-deploy.yaml`

## How tests are built from directories

As noted in the previous section, most tests are not defined in a single yaml file, but rather as a combination of files collected from a directory tree within the `suites/` subdirectory of the `ceph/qa sub-directory`.

The set of all tests defined by a given subdirectory of `suites/` is called an “integration test suite”, or a “teuthology suite”.

Combination of yaml facets is controlled by special files (`%` and `+`) that are placed within the directory tree and can be thought of as operators. The `%` file is the “convolution” operator and `+` signifies concatenation.

## Convolution operator

The convolution operator, implemented as an empty file called `%`, tells teuthology to construct a test matrix from yaml facets found in subdirectories below the directory containing the operator.

For example, the `ceph-deploy suite` is defined by the `suites/ceph-deploy/` tree, which consists of the files and subdirectories in the following structure

```
1. qa/suites/ceph-deploy
2.   └── %
3.   └── distros
```

```

4. |   └── centos_latest.yaml
5. |   └── ubuntu_latest.yaml
6. └── tasks
7.     ├── ceph-admin-commands.yaml
8.     └── rbd_import_export.yaml

```

This is interpreted as a 2x1 matrix consisting of two tests:

1. ceph-deploy/basic/{distros/centos\_7.0.yaml tasks/ceph-deploy.yaml}
2. ceph-deploy/basic/{distros/ubuntu\_16.04.yaml tasks/ceph-deploy.yaml}

i.e. the concatenation of centos\_7.0.yaml and ceph-deploy.yaml and the concatenation of ubuntu\_16.04.yaml and ceph-deploy.yaml, respectively. In human terms, this means that the task found in `ceph-deploy.yaml` is intended to run on both CentOS 7.0 and Ubuntu 16.04.

Without the file percent, the `ceph-deploy` tree would be interpreted as three standalone tests:

- ceph-deploy/basic/distros/centos\_7.0.yaml
- ceph-deploy/basic/distros/ubuntu\_16.04.yaml
- ceph-deploy/basic/tasks/ceph-deploy.yaml

(which would of course be wrong in this case).

Referring to the `ceph/qa` sub-directory, you will notice that the `centos_7.0.yaml` and `ubuntu_16.04.yaml` files in the `suites/ceph-deploy/basic/distros/` directory are implemented as symlinks. By using symlinks instead of copying, a single file can appear in multiple suites. This eases the maintenance of the test framework as a whole.

All the tests generated from the `suites/ceph-deploy/` directory tree (also known as the “ceph-deploy suite”) can be run with

```
1. teuthology-suite --machine-type smithi --suite ceph-deploy
```

An individual test from the `ceph-deploy suite` can be run by adding the `--filter` option

```

1. teuthology-suite \
2.   --machine-type smithi \
3.   --suite ceph-deploy/basic \
4.   --filter 'ceph-deploy/basic/{distros/ubuntu_16.04.yaml tasks/ceph-deploy.yaml}'

```

## Note

To run a standalone test like the one in [Reading a standalone test](#), `--suite` alone is sufficient. If you want to run a single test from a suite that is defined as a

directory tree, `--suite` must be combined with `--filter`. This is because the `--suite` option understands POSIX relative paths only.

## Concatenation operator

For even greater flexibility in sharing yaml files between suites, the special file plus (`+`) can be used to concatenate files within a directory. For instance, consider the `suites/rbd/thrash` tree

```

1. qa/suites/rbd/thrash
2.   └── %
3.     ├── clusters
4.     |   └── +
5.     |       └── fixed-2.yaml
6.     |       └── openstack.yaml
7.     └── workloads
8.         ├── rbd_api_tests_copy_on_read.yaml
9.         └── rbd_api_tests.yaml
10.            └── rbd_fsx_rate_limit.yaml

```

This creates two tests:

- `rbd/thrash/{clusters/fixed-2.yaml clusters/openstack.yaml workloads/rbd_api_tests_copy_on_read.yaml}`
- `rbd/thrash/{clusters/fixed-2.yaml clusters/openstack.yaml workloads/rbd_api_tests.yaml}`

Because the `clusters/` subdirectory contains the special file plus (`+`), all the other files in that subdirectory (`fixed-2.yaml` and `openstack.yaml` in this case) are concatenated together and treated as a single file. Without the special file plus, they would have been convolved with the files from the `workloads` directory to create a  $2 \times 2$  matrix:

- `rbd/thrash/{clusters/openstack.yaml workloads/rbd_api_tests_copy_on_read.yaml}`
- `rbd/thrash/{clusters/openstack.yaml workloads/rbd_api_tests.yaml}`
- `rbd/thrash/{clusters/fixed-2.yaml workloads/rbd_api_tests_copy_on_read.yaml}`
- `rbd/thrash/{clusters/fixed-2.yaml workloads/rbd_api_tests.yaml}`

The `clusters/fixed-2.yaml` file is shared among many suites to define the following `roles`

```

1. roles:
2. - [mon.a, mon.c, osd.0, osd.1, osd.2, client.0]
3. - [mon.b, osd.3, osd.4, osd.5, client.1]

```

The `rbd/thrash` suite as defined above, consisting of two tests, can be run with

```
1. teuthology-suite --machine-type smithi --suite rbd/thrash
```

A single test from the rbd/thrash suite can be run by adding the `--filter` option

```
1. teuthology-suite \
2.   --machine-type smithi \
3.   --suite rbd/thrash \
4.   --filter 'rbd/thrash/{clusters/fixed-2.yaml clusters/openstack.yaml
4. workloads/rbd_api_tests_copy_on_read.yaml}'
```

## Filtering tests by their description

When a few jobs fail and need to be run again, the `--filter` option can be used to select tests with a matching description. For instance, if the `rados` suite fails the `all/peer.yaml` test, the following will only run the tests that contain this file

```
1. teuthology-suite --machine-type smithi --suite rados --filter all/peer.yaml
```

The `--filter-out` option does the opposite (it matches tests that do not contain a given string), and can be combined with the `--filter` option.

Both `--filter` and `--filter-out` take a comma-separated list of strings (which means the comma character is implicitly forbidden in filenames found in the [ceph/qa sub-directory](#)). For instance

```
1. teuthology-suite --machine-type smithi --suite rados --filter all/peer.yaml,all/rest-api.yaml
```

will run tests that contain either `all/peer.yaml` or `all/rest-api.yaml`

Each string is looked up anywhere in the test description and has to be an exact match: they are not regular expressions.

## Reducing the number of tests

The `rados` suite generates tens or even hundreds of thousands of tests out of a few hundred files. This happens because teuthology constructs test matrices from subdirectories wherever it encounters a file named `%`. For instance, all tests in the `rados/basic suite` run with different messenger types: `simple`, `async` and `random`, because they are combined (via the special file `%`) with the `msgr directory`

All integration tests are required to be run before a Ceph release is published. When merely verifying whether a contribution can be merged without risking a trivial regression, it is enough to run a subset. The `--subset` option can be used to reduce the number of tests that are triggered. For instance

```
1. teuthology-suite --machine-type smithi --suite rados --subset 0/4000
```

will run as few tests as possible. The tradeoff in this case is that not all combinations of test variations will together, but no matter how small a ratio is provided in the `--subset`, teuthology will still ensure that all files in the suite are in at least one test. Understanding the actual logic that drives this requires reading the teuthology source code.

The `--limit` option only runs the first `N` tests in the suite: this is rarely useful, however, because there is no way to control which test will be first.

# Running Unit Tests

## How to run s3-tests locally

RGW code can be tested by building Ceph locally from source, starting a vstart cluster, and running the “s3-tests” suite against it.

The following instructions should work on jewel and above.

### Step 1 - build Ceph

Refer to [Build Ceph](#).

You can do step 2 separately while it is building.

### Step 2 - vstart

When the build completes, and still in the top-level directory of the git clone where you built Ceph, do the following, for cmake builds:

```
1. cd build/  
2. RGW=1 ./src/vstart.sh -n
```

This will produce a lot of output as the vstart cluster is started up. At the end you should see a message like:

```
1. started. stop.sh to stop. see out/* (e.g. 'tail -f out/????') for debug output.
```

This means the cluster is running.

### Step 3 - run s3-tests

To run the s3tests suite do the following:

```
1. $ ./qa/workunits/rgw/run-s3tests.sh
```

## Running test using vstart\_runner.py

CephFS and Ceph Manager code is be tested using [vstart\\_runner.py](#).

## Running your first test

The Python tests in Ceph repository can be executed on your local machine using [vstart\\_runner.py](#). To do that, you'd need [teuthology](#) installed:

```
1. $ virtualenv --python=python3 venv
2. $ source venv/bin/activate
3. $ pip install 'setuptools >= 12'
4. $ pip install git+https://github.com/ceph/teuthology#egg=teuthology[test]
5. $ deactivate
```

The above steps installs teuthology in a virtual environment. Before running a test locally, build Ceph successfully from the source (refer [Build Ceph](#)) and do:

```
1. $ cd build
2. $ ./src/vstart.sh -n -d -l
3. $ source ~/path/to/teuthology/venv/bin/activate
```

To run a specific test, say [test\\_reconnect\\_timeout](#) from [TestClientRecovery](#) in [qa/tasks/cephfs/test\\_client\\_recovery](#), you can do:

```
$ python ../qa/tasks/vstart_runner.py
1. tasks.cephfs.test_client_recovery.TestClientRecovery.test_reconnect_timeout
```

The above command runs [vstart\\_runner.py](#) and passes the test to be executed as an argument to [vstart\\_runner.py](#). In a similar way, you can also run the group of tests in the following manner:

```
1. $ # run all tests in class TestClientRecovery
2. $ python ../qa/tasks/vstart_runner.py tasks.cephfs.test_client_recovery.TestClientRecovery
3. $ # run all tests in test_client_recovery.py
4. $ python ../qa/tasks/vstart_runner.py tasks.cephfs.test_client_recovery
```

Based on the argument passed, [vstart\\_runner.py](#) collects tests and executes as it would execute a single test.

[vstart\\_runner.py](#) can take the following options -

--clear-old-log

deletes old log file before running the test

--create

create Ceph cluster before running a test

--create-cluster-only

creates the cluster and quits; tests can be issued later

--interactive

```

drops a Python shell when a test fails

--log-ps-output

logs ps output; might be useful while debugging

--teardown

tears Ceph cluster down after test(s) has finished runnng

--kclient

use the kernel cephfs client instead of FUSE

--brxnet=<net/mask>

specify a new net/mask for the mount clients' network namespace container (Default: 192.168.0.0/16)

```

#### Note

If using the FUSE client, ensure that the fuse package is installed and enabled on the system and that `user_allow_other` is added to `/etc/fuse.conf`.

#### Note

If using the kernel client, the user must have the ability to run commands with passwordless sudo access. A failure on the kernel client may crash the host, so it's recommended to use this functionality within a virtual machine.

## Internal working of vstart\_runner.py -

vstart\_runner.py primarily does three things -

- collects and runs the tests

vstart\_runner.py setups/teardowns the cluster and collects and runs the test. This is implemented using methods `scan_tests()`, `load_tests()` and `exec_test()`. This is where all the options that vstart\_runner.py takes are implemented along with other features like logging and copying the traceback to the bottom of the log.

- provides an interface for issuing and testing shell commands

The tests are written assuming that the cluster exists on remote machines. vstart\_runner.py provides an interface to run the same tests with the cluster that exists within the local machine. This is done using the class `LocalRemote`. Class `LocalRemoteProcess` can manage the process that executes the commands from `LocalRemote`, class `LocalDaemon` provides an interface to handle Ceph daemons and class `LocalFuseMount` can create and handle FUSE mounts.

- provides an interface to operate Ceph cluster

`LocalCephManager` provides methods to run Ceph cluster commands with and without admin socket and `LocalCephCluster` provides methods to set or clear `ceph.conf`.

# Running Integration Tests using Teuthology

## Getting binaries

To run integration tests using teuthology, you need to have Ceph binaries built for your branch. Follow these steps to initiate the build process -

1. Push the branch to `ceph-ci` repository. This triggers the process of building the binaries.
2. To confirm that the build process has been initiated, spot the branch name at [Shaman](#). Little after the build process has been initiated, the single entry with your branch name would multiply, each new entry for a different combination of distro and flavour.
3. Wait until the packages are built and uploaded, and the repository offering them are created. This is marked by colouring the entries for the branch name green. Preferably, wait until each entry is coloured green. Usually, it takes around 2-3 hours depending on the availability of the machines.

### Note

Branch to be pushed on `ceph-ci` can be any branch, it shouldn't necessarily be a PR branch.

### Note

In case you are pushing master or any other standard branch, check [Shaman](#) beforehand since it already might have builds ready for it.

## Triggering Tests

After building is complete, proceed to trigger tests -

1. Log in to the teuthology machine:

```
1. ssh <username>@teuthology.front.sepia.ceph.com
```

This would require Sepia lab access. To know how to request it, see:  
[https://ceph.github.io/sepia/adding\\_users/](https://ceph.github.io/sepia/adding_users/)

2. Next, get teuthology installed. Run the first set of commands in [Running Your First Test](#) for that. After that, activate the virtual environment in which teuthology is installed.
3. Run the `teuthology-suite` command:

```
1. teuthology-suite -v -m smithi -c wip-devname-feature-x -s fs -p 110 --filter "cephfs-shell"
```

Following are the options used in above command with their meanings -

- -v  
verbose
- m  
machine name
- c  
branch name, the branch that was pushed on ceph-ci
- s  
test-suite name
- p  
higher the number, lower the priority of the job
- filter  
filter tests in given suite that needs to run, the arg to filter should be the test you want to run

#### Note

The priority number present in the command above is just a placeholder. It might be highly inappropriate for the jobs you may want to trigger. See [Testing Priority](#) section to pick a priority number.

#### Note

Don't skip passing a priority number, the default value is 1000 which way too high; the job probably might never run.

1. Wait for the tests to run. `teuthology-suite` prints a link to the [Pulpito](#) page created for the tests triggered.

Other frequently used/useful options are `-d` (or `--distro`), `--distroversion`, `--filter-out`, `--timeout`, `flavor`, `-rerun`, `-l` (for limiting number of jobs), `-n` (for how many times job would run) and `-e` (for email notifications). Run `teuthology-suite --help` to read description of these and every other options available.

## Testing QA changes (without re-building binaries)

While writing a PR you might need to test your PR repeatedly using teuthology. If you are making non-QA changes, you need to follow the standard process of triggering builds, waiting for it to finish and then triggering tests and wait for the result. But if changes you made are purely changes in qa/, you don't need rebuild the binaries. Instead you can test binaries built for the ceph-ci branch and instruct

`teuthology-suite` command to use a separate branch for running tests. The separate branch can be passed to the command by using `--suite-repo` and `--suite-branch`. Pass the link to the GitHub fork where your PR branch exists to the first option and pass the PR branch name to the second option.

For example, if you want to make changes in `qa/` after testing `branch-x` (of which has ceph-ci branch is `wip-username-branch-x`) by running following command:

```
1. teuthology-suite -v -m smithi -c wip-username-branch-x -s fs -p 50 --filter cephfs-shell
```

You can make the modifications locally, update the PR branch and then trigger tests from your PR branch as follows:

```
teuthology-suite -v -m smithi -c wip-username-branch-x -s fs -p 50 --filter cephfs-shell --suite-repo  
1. https://github.com/username/ceph --suite-branch branch-x
```

You can verify if the tests were run using this branch by looking at values for the keys `suite_branch`, `suite_repo` and `suite_sha1` in the job config printed at the very beginning of the teuthology job.

## About Suites and Filters

See [Suites Inventory](#) for a list of suites of integration tests present right now. Alternatively, each directory under `qa/suites` in Ceph repository is an integration test suite, so looking within that directory to decide an appropriate argument for `-s` also works.

For picking an argument for `--filter`, look within `qa/suites/<suite-name>/<subsuite-name>/tasks` to get keywords for filtering tests. Each YAML file in there can trigger a bunch of tests; using the name of the file, without the extension part of the file name, as an argument to the `--filter` will trigger those tests. For example, the sample command above uses `cephfs-shell` since there's a file named `cephfs-shell.yaml` in `qa/suites/fs/basic_functional/tasks/`. In case, the file name doesn't hint what bunch of tests it would trigger, look at the contents of the file for `modules` attribute. For `cephfs-shell.yaml` the `modules` attribute is `tasks.cephfs.test_cephfs_shell` which means it'll trigger all tests in `qa/tasks/cephfs/test_cephfs_shell.py`.

## Killing Tests

Sometimes a teuthology job might not complete running for several minutes or even hours after tests that were triggered have completed running and other times wrong set

of tests can be triggered is filter wasn't chosen carefully. To save resource it's better to terminate such a job. Following is the command to terminate a job:

```
1. teuthology-kill -r teuthology-2019-12-10_05:00:03-smoke-master-testing-basic-smithi
```

Let's call the argument passed to `-r` as test ID. It can be found easily in the link to the Pulpito page for the tests you triggered. For example, for the above test ID, the link is - [http://pulpito.front.sepia.ceph.com/teuthology-2019-12-10\\_05:00:03-smoke-master-testing-basic-smithi/](http://pulpito.front.sepia.ceph.com/teuthology-2019-12-10_05:00:03-smoke-master-testing-basic-smithi/)

## Re-running Tests

Pass `--rerun` option, with test ID as an argument to it, to `teuthology-suite` command:

```
teuthology-suite -v -m smithi -c wip-rishabh-fs-test_cephfs_shell-fix -p 50 --rerun teuthology-2019-12-10_05:00:03-smoke-master-testing-basic-smithi
```

The meaning of rest of the options is already covered in Triggering Tests section.

## Teuthology Archives

Once the tests have finished running, the log for the job can be obtained by clicking on job ID at the Pulpito page for your tests. It's more convenient to download the log and then view it rather than viewing it in an internet browser since these logs can easily be upto size of 1 GB. What's much more easier is to log in to the teuthology machine again (`teuthology.front.sepia.ceph.com`), and access the following path:

```
1. /ceph/teuthology-archive/<test-id>/<job-id>/teuthology.log
```

For example, for above test ID path is:

```
1. /ceph/teuthology-archive/teuthology-2019-12-10_05:00:03-smoke-master-testing-basic-smithi/4588482/teuthology.log
```

This way the log remotely can be viewed remotely without having to wait too much.

## Naming the ceph-ci branch

There are no hard conventions (except for the case of stable branch; see next paragraph) for how the branch pushed on ceph-ci is named. But, to make builds and tests easily identifiable on Shaman and Pulpito respectively, prepend it with your name. For example branch `feature-x` can be named `wip-yourname-feature-x` while pushing on ceph-ci.

In case you are using one of the stable branches (e.g. nautilus, mimic, etc.), include

the name of that stable branch in your ceph-ci branch name. For example, `feature-x` PR branch should be named as `wip-feature-x-nautilus`. *This is not just a matter of convention but this, more essentially, builds your branch in the correct environment.*

Delete the branch from ceph-ci, once it's not required anymore. If you are logged in at GitHub, all your branches on ceph-ci can be easily found here - <https://github.com/ceph/ceph-ci/branches>.

# Running Tests in the Cloud

In this chapter, we will explain in detail how use an OpenStack tenant as an environment for Ceph [integration testing](#).

## Assumptions and caveat

We assume that:

1. you are the only person using the tenant
2. you have the credentials
3. the tenant supports the `nova` and `cinder` APIs

Caveat: be aware that, as of this writing (July 2016), testing in OpenStack clouds is a new feature. Things may not work as advertised. If you run into trouble, ask for help on [IRC](#) or the [Mailing list](#), or open a bug report at the [ceph-workbench bug tracker](#).

## Prepare tenant

If you have not tried to use `ceph-workbench` with this tenant before, proceed to the next step.

To start with a clean slate, login to your tenant via the Horizon dashboard and:

- terminate the `teuthology` and `packages-repository` instances, if any
- delete the `teuthology` and `teuthology-worker` security groups, if any
- delete the `teuthology` and `teuthology-myself` key pairs, if any

Also do the above if you ever get key-related errors (“invalid key”, etc.) when trying to schedule suites.

## Getting ceph-workbench

Since testing in the cloud is done using the `ceph-workbench ceph-qa-suite` tool, you will need to install that first. It is designed to be installed via Docker, so if you don’t have Docker running on your development machine, take care of that first. You can follow [the official tutorial](#) to install if you have not installed yet.

Once Docker is up and running, install `ceph-workbench` by following the [Installation instructions in the ceph-workbench documentation](#).

# Linking ceph-workbench with your OpenStack tenant

Before you can trigger your first teuthology suite, you will need to link `ceph-workbench` with your OpenStack account.

First, download a `openrc.sh` file by clicking on the “Download OpenStack RC File” button, which can be found in the “API Access” tab of the “Access & Security” dialog of the OpenStack Horizon dashboard.

Second, create a `~/.ceph-workbench` directory, set its permissions to 700, and move the `openrc.sh` file into it. Make sure that the filename is exactly `~/.ceph-workbench/openrc.sh`.

Third, edit the file so it does not ask for your OpenStack password interactively. Comment out the relevant lines and replace them with something like:

```
1. ... prompt:: bash $  
  
export OS_PASSWORD="aiVeth0aejee3eep8rogho3eep7Pha6ek"
```

When `ceph-workbench ceph-qa-suite` connects to your OpenStack tenant for the first time, it will generate two keypairs: `teuthology-myself` and `teuthology`.

## Run the dummy suite

You are now ready to take your OpenStack teuthology setup for a test drive

```
1. ceph-workbench ceph-qa-suite --suite dummy
```

Be forewarned that the first run of `ceph-workbench ceph-qa-suite` on a pristine tenant will take a long time to complete because it downloads a VM image and during this time the command may not produce any output.

The images are cached in OpenStack, so they are only downloaded once. Subsequent runs of the same command will complete faster.

Although `dummy` suite does not run any tests, in all other respects it behaves just like a teuthology suite and produces some of the same artifacts.

The last bit of output should look something like this:

```
1. pulpito web interface: http://149.202.168.201:8081/  
ssh access : ssh -i /home/smithfarm/.ceph-workbench/teuthology-myself.pem ubuntu@149.202.168.201 #  
2. logs in /usr/share/nginx/html
```

What this means is that `ceph-workbench ceph-qa-suite` triggered the test suite run. It does

not mean that the suite run has completed. To monitor progress of the run, check the Pulpito web interface URL periodically, or if you are impatient, ssh to the teuthology machine using the ssh command shown and do

```
1. tail -f /var/log/teuthology.*
```

The /usr/share/nginx/html directory contains the complete logs of the test suite. If we had provided the `--upload` option to the `ceph-workbench ceph-qa-suite` command, these logs would have been uploaded to <http://teuthology-logs.public.ceph.com>.

## Run a standalone test

The standalone test explained in [Reading a standalone test](#) can be run with the following command

```
1. ceph-workbench ceph-qa-suite --suite rados/singleton/all/admin-socket.yaml
```

This will run the suite shown on the current `master` branch of `ceph/ceph.git`. You can specify a different branch with the `--ceph` option, and even a different git repo with the `--ceph-git-url` option. (Run `ceph-workbench ceph-qa-suite --help` for an up-to-date list of available options.)

The first run of a suite will also take a long time, because ceph packages have to be built, first. Again, the packages so built are cached and `ceph-workbench ceph-qa-suite` will not build identical packages a second time.

## Interrupt a running suite

Teuthology suites take time to run. From time to time one may wish to interrupt a running suite. One obvious way to do this is:

```
1. ... prompt:: bash $
```

```
ceph-workbench ceph-qa-suite -teardown
```

This destroys all VMs created by `ceph-workbench ceph-qa-suite` and returns the OpenStack tenant to a “clean slate”.

Sometimes you may wish to interrupt the running suite, but keep the logs, the teuthology VM, the packages-repository VM, etc. To do this, you can `ssh` to the teuthology VM (using the `ssh access` command reported when you triggered the suite – see [Run the dummy suite](#)) and, once there

```
1. sudo /etc/init.d/teuthology restart
```

This will keep the teuthology machine, the logs and the packages-repository instance but nuke everything else.

## Upload logs to archive server

Since the teuthology instance in OpenStack is only semi-permanent, with limited space for storing logs, `teuthology-openstack` provides an `--upload` option which, if included in the `ceph-workbench ceph-qa-suite` command, will cause logs from all failed jobs to be uploaded to the log archive server maintained by the Ceph project. The logs will appear at the URL:

```
1. http://teuthology-logs.public.ceph.com/$RUN
```

where `$RUN` is the name of the run. It will be a string like this:

```
1. ubuntu-2016-07-23_16:08:12-rados-hammer-backports---basic-openstack
```

Even if you don't providing the `--upload` option, however, all the logs can still be found on the teuthology machine in the directory `/usr/share/nginx/html`.

## Provision VMs ad hoc

From the teuthology VM, it is possible to provision machines on an "ad hoc" basis, to use however you like. The magic incantation is:

```
1. ... prompt:: bash $
```

```
teuthology-lock -lock-many $NUMBER_OF_MACHINES  
-os-type $OPERATING_SYSTEM -os-version $OS_VERSION -machine-type openstack -owner $EMAIL_ADDRESS
```

The command must be issued from the `~/teuthology` directory. The possible values for `OPERATING_SYSTEM` AND `OS_VERSION` can be found by examining the contents of the directory `teuthology/openstack/`. For example

```
1. teuthology-lock --lock-many 1 --os-type ubuntu --os-version 16.04 \  
2. --machine-type openstack --owner foo@example.com
```

When you are finished with the machine, find it in the list of machines

```
1. openstack server list
```

to determine the name or ID, and then terminate it with

```
1. openstack server delete $NAME_OR_ID
```

# Deploy a cluster for manual testing

The [teuthology framework](#) and [ceph-workbench ceph-qa-suite](#) are versatile tools that automatically provision Ceph clusters in the cloud and run various tests on them in an automated fashion. This enables a single engineer, in a matter of hours, to perform thousands of tests that would keep dozens of human testers occupied for days or weeks if conducted manually.

However, there are times when the automated tests do not cover a particular scenario and manual testing is desired. It turns out that it is simple to adapt a test to stop and wait after the Ceph installation phase, and the engineer can then ssh into the running cluster. Simply add the following snippet in the desired place within the test YAML and schedule a run with the test:

```
1. tasks:
2. - exec:
3.   client.0:
4.     - sleep 1000000000 # forever
```

(Make sure you have a `client.0` defined in your `roles` stanza or adapt accordingly.)

The same effect can be achieved using the `interactive` task:

```
1. tasks:
2. - interactive
```

By following the test log, you can determine when the test cluster has entered the “sleep forever” condition. At that point, you can ssh to the teuthology machine and from there to one of the target VMs (OpenStack) or teuthology worker machines machine (Sepia) where the test cluster is running.

The VMs (or “instances” in OpenStack terminology) created by [ceph-workbench ceph-qa-suite](#) are named as follows:

- `teuthology` - the teuthology machine
- `packages-repository` - VM where packages are stored
- `ceph-*` - VM where packages are built
- `target*` - machines where tests are run

The VMs named `target*` are used by tests. If you are monitoring the teuthology log for a given test, the hostnames of these target machines can be found out by searching for the string `Locked targets`:

```
1. 2016-03-20T11:39:06.166 INFO:teuthology.task.internal:Locked targets:
```

```
2. target149202171058.teuthology: null  
3. target149202171059.teuthology: null
```

The IP addresses of the target machines can be found by running `openstack server list` on the teuthology machine, but the target VM hostnames (e.g. `target149202171058.teuthology`) are resolvable within the teuthology cluster.

# Ceph Dashboard Developer Documentation

---

## Table of Contents

- Ceph Dashboard Developer Documentation
  - Feature Design
  - Preliminary Steps
    - vstart
    - Host-based vs Docker-based Development Environments
      - Development environment on your host system
      - Development environments based on Docker
    - vstart on your host system
    - Docker
  - Frontend Development
    - Prerequisites
    - Package installation
    - Adding or updating packages
    - Setting up a Development Server
    - Code Scaffolding
    - Build the Project
    - Build the Code Documentation
    - Code linting and formatting
    - Ceph Dashboard and Bootstrap
    - Writing Unit Tests
    - Running Unit Tests
    - Running End-to-End (E2E) Tests
      - E2E Prerequisites
      - run-frontend-e2e-tests.sh

- [Other running options](#)
- [CYPRESS\\_CACHE\\_FOLDER](#)
- [Writing End-to-End Tests](#)
  - [The PagerHelper class](#)
  - [Subclasses of PageHelper](#)
    - [Helper Methods](#)
    - [Using PageHelpers](#)
  - [Code Style](#)
    - `describe()` vs `it()`
- [Differences between Frontend Unit Tests and End-to-End \(E2E\) Tests / FAQ](#)
  - [What are E2E/unit tests designed for?](#)
  - [Which E2E/unit tests are considered to be valid?](#)
  - [How should an E2E/unit test look like?](#)
  - [What should an E2E/unit test cover?](#)
  - [What should an E2E/unit test NOT cover?](#)
  - [Best practices/guideline](#)
- [Further Help](#)
- [Example of a Generator](#)
- [Frontend Typescript Code Style Guide Recommendations](#)
- [Frontend components](#)
- [Helper](#)
- [Terminology and wording](#)
- [Frontend branding](#)
- [UI Style Guide](#)
  - [Colors](#)
  - [Buttons](#)
  - [Links](#)

- [Forms](#)
- [Modals](#)
- [Icons](#)
- [Navigation](#)
- [Alerts and notifications](#)
- [I18N](#)
  - [How to extract messages from source code?](#)
  - [Supported languages](#)
  - [Translating process](#)
  - [Updating translated messages](#)
  - [Suggestions](#)
- [Backend Development](#)
  - [Unit Testing](#)
  - [Unit tests based on tox](#)
  - [API tests based on Teuthology](#)
  - [How to add a new controller?](#)
  - [Implementing Proxy Controller](#)
  - [How does the RESTController work?](#)
  - [How to use a custom API endpoint in a RESTController?](#)
  - [How to restrict access to a controller?](#)
  - [How to create a dedicated UI endpoint which uses the 'public' API?](#)
  - [How to access the manager module instance from a controller?](#)
  - [How to write a unit test for a controller?](#)
  - [How to listen for manager notifications in a controller?](#)
  - [How to write a unit test when a controller accesses a Ceph module?](#)
  - [How to add a new configuration setting?](#)
  - [How to run a controller read-write operation asynchronously?](#)

- How to get the list of executing and finished asynchronous tasks?
- How to use asynchronous APIs with asynchronous tasks?
- How to update the execution progress of an asynchronous task?
- How to deal with asynchronous tasks in the front-end?
- REST API documentation
- Error Handling in Python
- Plug-ins

## Feature Design

---

To promote collaboration on new Ceph Dashboard features, the first step is the definition of a design document. These documents then form the basis of implementation scope and permit wider participation in the evolution of the Ceph Dashboard UI.

Design Documents:

- [UI Design Goals](#)

## Preliminary Steps

---

The following documentation chapters expect a running Ceph cluster and at least a running `dashboard` manager module (with few exceptions). This chapter gives an introduction on how to set up such a system for development, without the need to set up a full-blown production environment. All options introduced in this chapter are based on a so called `vstart` environment.

### Note

Every `vstart` environment needs Ceph [to be compiled](#) from its Github repository, though Docker environments simplify that step by providing a shell script that contains those instructions.

One exception to this rule are the [build-free](#) capabilities of `ceph-dev`. See below for more information.

## `vstart`

`"vstart"` is actually a shell script in the `src/` directory of the Ceph repository (`src/vstart.sh`). It is used to start a single node Ceph cluster on the machine where it is executed. Several required and some optional Ceph internal services are started automatically when it is used to start a Ceph cluster. `vstart` is the basis for the three most commonly used development environments in Ceph Dashboard.

You can read more about vstart in [Deploying a development cluster](#). Additional information for developers can also be found in the [Developer Guide](#).

## Host-based vs Docker-based Development Environments

This document introduces you to three different development environments, all based on vstart. Those are:

- vstart running on your host system
- vstart running in a Docker environment
  - [ceph-dev-docker](#)
  - [ceph-dev](#)

Besides their independent development branches and sometimes slightly different approaches, they also differ with respect to their underlying operating systems.

Release	ceph-dev-docker	ceph-dev
Mimic	openSUSE Leap 15	CentOS 7
Nautilus	openSUSE Leap 15	CentOS 7
Octopus	openSUSE Leap 15.2	CentOS 8
Master	openSUSE Tumbleweed	CentOS 8

### Note

Independently of which of these environments you will choose, you need to compile Ceph in that environment. If you compiled Ceph on your host system, you would have to recompile it on Docker to be able to switch to a Docker based solution. The same is true vice versa. If you previously used a Docker development environment and compiled Ceph there and you now want to switch to your host system, you will also need to recompile Ceph (or compile Ceph using another separate repository).

[ceph-dev](#) is an exception to this rule as one of the options it provides is [build-free](#). This is accomplished through a Ceph installation using RPM system packages. You will still be able to work with a local Github repository like you are used to.

## Development environment on your host system

- No need to learn or have experience with Docker, jump in right away.
- Limited amount of scripts to support automation (like Ceph compilation).
- No pre-configured easy-to-start services (Prometheus, Grafana, etc).

- Limited amount of host operating systems supported, depending on which Ceph version is supposed to be used.
- Dependencies need to be installed on your host.
- You might find yourself in the situation where you need to upgrade your host operating system (for instance due to a change of the GCC version used to compile Ceph).

## Development environments based on Docker

- Some overhead in learning Docker if you are not used to it yet.
- Both Docker projects provide you with scripts that help you getting started and automate recurring tasks.
- Both Docker environments come with partly pre-configured external services which can be used to attach to or complement Ceph Dashboard features, like
  - Prometheus
  - Grafana
  - Node-Exporter
  - Shibboleth
  - HAProxy
- Works independently of the operating system you use on your host.

## vstart on your host system

The vstart script is usually called from your build/ directory like so:

```
1. .../src/vstart.sh -n -d
```

In this case `-n` ensures that a new vstart cluster is created and that a possibly previously created cluster isn't re-used. `-d` enables debug messages in log files. There are several more options to chose from. You can get a list using the `--help` argument.

At the end of the output of vstart, there should be information about the dashboard and its URLs:

```
1. vstart cluster complete. Use stop.sh to stop. See out/* (e.g. 'tail -f out/????') for debug output.
2.
3. dashboard urls: https://192.168.178.84:41259, https://192.168.178.84:43259, https://192.168.178.84:45259
4. w/ user/pass: admin / admin
5. restful urls: https://192.168.178.84:42259, https://192.168.178.84:44259, https://192.168.178.84:46259
```

```
6. w/ user/pass: admin / 598da51f-8cd1-4161-a970-b2944d5ad200
```

During development (especially in backend development), you also want to check on occasions if the dashboard manager module is still running. To do so you can call `./bin/ceph mgr services` manually. It will list all the URLs of successfully enabled services. Only URLs of services which are available over HTTP(S) will be listed there. Ceph Dashboard is one of these services. It should look similar to the following output:

```
1. $ ./bin/ceph mgr services
2. {
3.     "dashboard": "https://home:41931/",
4.     "restful": "https://home:42931/"
5. }
```

By default, this environment uses a randomly chosen port for Ceph Dashboard and you need to use this command to find out which one it has become.

## Docker

Docker development environments usually ship with a lot of useful scripts. `ceph-dev-docker` for instance contains a file called `start-ceph.sh`, which cleans up log files, always starts a Rados Gateway service, sets some Ceph Dashboard configuration options and automatically runs a frontend proxy, all before or after starting up your `vstart` cluster.

Instructions on how to use those environments are contained in their respective repository README files.

- [ceph-dev-docker](#)
- [ceph-dev](#)

## Frontend Development

Before you can start the dashboard from within a development environment, you will need to generate the frontend code and either use a compiled and running Ceph cluster (e.g. started by `vstart.sh`) or the standalone development web server.

The build process is based on [Node.js](#) and requires the [Node Package Manager](#) `npm` to be installed.

## Prerequisites

- Node 10.0.0 or higher
- NPM 5.7.0 or higher

## nodeenv:

During Ceph's build we create a virtualenv with `node` and `npm` installed, which can be used as an alternative to installing node/npm in your system.

If you want to use the node installed in the virtualenv you just need to activate the virtualenv before you run any npm commands. To activate it run `.`

```
build/src/pybind/mgr/dashboard/node-env/bin/activate .
```

Once you finish, you can simply run `deactivate` and exit the virtualenv.

## Angular CLI:

If you do not have the [Angular CLI](#) installed globally, then you need to execute `ng` commands with an additional `npm run` before it.

## Package installation

Run `npm ci` in directory `src/pybind/mgr/dashboard/frontend` to install the required packages locally.

## Adding or updating packages

Run the following commands to add/update a package:

1. `npm install <PACKAGE_NAME>`
2. `npm run fix:audit`
3. `npm ci`

`fix:audit` is required because we have some packages that need to be fixed to a specific version and `npm install` tends to overwrite this.

## Setting up a Development Server

Create the `proxy.conf.json` file based on `proxy.conf.json.sample`.

Run `npm start` for a dev server. Navigate to `http://localhost:4200/`. The app will automatically reload if you change any of the source files.

## Code Scaffolding

Run `ng generate component component-name` to generate a new component. You can also use `ng generate directive|pipe|service|class|guard|interface|enum|module`.

## Build the Project

Run `npm run build` to build the project. The build artifacts will be stored in the `dist/` directory. Use the `--prod` flag for a production build (`npm run build -- --prod`). Navigate to <https://localhost:8443>.

## Build the Code Documentation

Run `npm run doc-build` to generate code docs in the `documentation/` directory. To make them accessible locally for a web browser, run `npm run doc-serve` and they will become available at <http://localhost:8444>. With `npm run compodoc -- <opts>` you may [fully configure it](#).

## Code linting and formatting

We use the following tools to lint and format the code in all our TS, SCSS and HTML files:

- [codelyzer](#)
- [html-linter](#)
- [htmllint-cli](#)
- [Prettier](#)
- [TSLint](#)
- [stylelint](#)

We added 2 npm scripts to help run these tools:

- `npm run lint`, will check frontend files against all linters
- `npm run fix`, will try to fix all the detected linting errors

## Ceph Dashboard and Bootstrap

Currently we are using Bootstrap on the Ceph Dashboard as a CSS framework. This means that most of our SCSS and HTML code can make use of all the utilities and other advantages Bootstrap is offering. In the past we often have used our own custom styles and this lead to more and more variables with a single use and double defined variables which sometimes are forgotten to be removed or it led to styling be inconsistent because people forgot to change a color or to adjust a custom SCSS class.

To get the current version of Bootstrap used inside Ceph please refer to the `package.json` and search for:

- `bootstrap` : For the Bootstrap version used.
- `@ng-bootstrap` : For the version of the Angular bindings which we are using.

So for the future please do the following when visiting a component:

- Does this HTML/SCSS code use custom code? - If yes: Is it needed? -> Clean it up before changing the things you want to fix or change.
- If you are creating a new component: Please make use of Bootstrap as much as reasonably possible! Don't try to reinvent the wheel.
- If possible please look up if Bootstrap has guidelines on how to extend it properly to do achieve what you want to achieve.

The more bootstrap alike our code is the easier it is to theme, to maintain and the less bugs we will have. Also since Bootstrap is a framework which tries to have usability and user experience in mind we increase both points exponentially. The biggest benefit of all is that there is less code for us to maintain which makes it easier to read for beginners and even more easy for people how are already familiar with the code.

## Writing Unit Tests

To write unit tests most efficient we have a small collection of tools, we use within test suites.

Those tools can be found under `src/pybind/mgr/dashboard/frontend/src/testing/`, especially take a look at `unit-test-helper.ts`.

There you will be able to find:

`configure TestBed` that replaces the initial `TestBed` methods. It takes the same arguments as `TestBed.configureTestingModuleTestingModule`. Using it will run your tests a lot faster in development, as it doesn't recreate everything from scratch on every test. To use the default behaviour pass `true` as the second argument.

`PermissionHelper` to help determine if the correct actions are shown based on the current permissions and selection in a list.

`FormHelper` which makes testing a form a lot easier with a few simple methods. It allows you to set a control or multiple controls, expect if a control is valid or has an error or just do both with one method. Additional you can expect a template element or multiple elements to be visible in the rendered template.

## Running Unit Tests

Run `npm run test` to execute the unit tests via `Jest`.

If you get errors on all tests, it could be because `Jest` or something else was updated. There are a few ways how you can try to resolve this:

- Remove all modules with `rm -rf dist node_modules` and run `npm install` again in order

to reinstall them

- Clear the cache of jest by running `npx jest --clearCache`

## Running End-to-End (E2E) Tests

We use [Cypress](#) to run our frontend E2E tests.

### E2E Prerequisites

You need to previously build the frontend.

In some environments, depending on your user permissions and the CYPRESS\_CACHE\_FOLDER, you might need to run `npm ci` with the `--unsafe-perm` flag.

You might need to install additional packages to be able to run Cypress. Please run `npx cypress verify` to verify it.

### `run-frontend-e2e-tests.sh`

Our `run-frontend-e2e-tests.sh` script is the go to solution when you wish to do a full scale e2e run. It will verify if everything needed is installed, start a new vstart cluster and run the full test suite.

Start all frontend E2E tests by running:

```
1. $ ./run-frontend-e2e-tests.sh
```

Report:

You can follow the e2e report on the terminal and you can find the screenshots of failed test cases by opening the following directory:

```
1. src/pybind/mgr/dashboard/frontend/cypress/screenshots/
```

Device:

You can force the script to use a specific device with the `-d` flag:

```
1. $ ./run-frontend-e2e-tests.sh -d <chrome|chromium|electron|docker>
```

Remote:

By default this script will stop and start a new vstart cluster. If you want to run the tests outside the ceph environment, you will need to manually define the dashboard url using `-r` and, optionally, credentials (`-u`, `-p`):

```
1. $ ./run-frontend-e2e-tests.sh -r <DASHBOARD_URL> -u <E2E_LOGIN_USER> -p <E2E_LOGIN_PWD>
```

Note:

When using docker, as your device, you might need to run the script with sudo permissions.

## Other running options

During active development, it is not recommended to run the previous script, as it is not prepared for constant file changes. Instead you should use one of the following commands:

- `npm run e2e` - This will run `ng serve` and open the Cypress Test Runner.
- `npm run e2e:ci` - This will run `ng serve` and run the Cypress Test Runner once.
- `npx cypress run` - This calls cypress directly and will run the Cypress Test Runner. You need to have a running frontend server.
- `npx cypress open` - This calls cypress directly and will open the Cypress Test Runner. You need to have a running frontend server.

Calling Cypress directly has the advantage that you can use any of the available [flags](#) to customize your test run and you don't need to start a frontend server each time.

Using one of the `open` commands, will open a cypress application where you can see all the test files you have and run each individually. This is going to be run in watch mode, so if you make any changes to test files, it will retrigger the test run. This cannot be used inside docker, as it requires X11 environment to be able to open.

By default Cypress will look for the web page at `https://localhost:4200/`. If you are serving it in a different URL you will need to configure it by exporting the environment variable `CYPRESS_BASE_URL` with the new value. E.g.:

```
CYPRESS_BASE_URL=https://localhost:41076/ npx cypress open
```

## CYPRESS\_CACHE\_FOLDER

When installing cypress via npm, a binary of the cypress app will also be downloaded and stored in a cache folder. This removes the need to download it every time you run `npm ci` or even when using cypress in a separate project.

By default Cypress uses `~/.cache` to store the binary. To prevent changes to the user home directory, we have changed this folder to `/ceph/build/src/pybind/mgr/dashboard/cypress`, so when you build ceph or run `run-frontend-e2e-tests.sh` this is the directory Cypress will use.

When using any other command to install or run cypress, it will go back to the default directory. It is recommended that you export the `CYPRESS_CACHE_FOLDER` environment variable with a fixed directory, so you always use the same directory no matter which command you use.

# Writing End-to-End Tests

## The PagerHelper class

The `PageHelper` class is supposed to be used for general purpose code that can be used on various pages or suites.

Examples are

- `navigateTo()` - Navigates to a specific page and waits for it to load
- `getFirstTableCell()` - returns the first table cell. You can also pass a string with the desired content and it will return the first cell that contains it.
- `getTabsCount()` - returns the amount of tabs

Every method that could be useful on several pages belongs there. Also, methods which enhance the derived classes of the `PageHelper` belong there. A good example for such a case is the `restrictTo()` decorator. It ensures that a method implemented in a subclass of `PageHelper` is called on the correct page. It will also show a developer-friendly warning if this is not the case.

## Subclasses of PageHelper

### Helper Methods

In order to make code reusable which is specific for a particular suite, make sure to put it in a derived class of the `PageHelper`. For instance, when talking about the pool suite, such methods would be `create()`, `exist()` and `delete()`. These methods are specific to a pool but are useful for other suites.

Methods that return HTML elements which can only be found on a specific page, should be either implemented in the helper methods of the subclass of `PageHelper` or as own methods of the subclass of `PageHelper`.

### Using PageHelpers

In any suite, an instance of the specific `Helper` class should be instantiated and called directly.

```

1. const pools = new PoolPageHelper();
2.
3. it('should create a pool', () => {
4.   pools.exist(poolName, false);
5.   pools.navigateTo('create');
6.   pools.create(poolName, 8);
7.   pools.exist(poolName, true);
8. });

```

## Code Style

Please refer to the official [Cypress Core Concepts](#) for a better insight on how to write and structure tests.

### `describe()` vs `it()`

Both `describe()` and `it()` are function blocks, meaning that any executable code necessary for the test can be contained in either block. However, Typescript scoping rules still apply, therefore any variables declared in a `describe` are available to the `it()` blocks inside of it.

`describe()` typically are containers for tests, allowing you to break tests into multiple parts. Likewise, any setup that must be made before your tests are run can be initialized within the `describe()` block. Here is an example:

```

1. describe('create, edit & delete image test', () => {
2.   const poolName = 'e2e_images_pool';
3.
4.   before(() => {
5.     cy.login();
6.     pools.navigateTo('create');
7.     pools.create(poolName, 8, 'rbd');
8.     pools.exist(poolName, true);
9.   });
10.
11. beforeEach(() => {
12.   cy.login();
13.   images.navigateTo();
14. });
15.
16. //...
17.
18. });

```

As shown, we can initiate the variable `poolName` as well as run commands before our test suite begins (creating a pool). `describe()` block messages should include what the test suite is.

`it()` blocks typically are parts of an overarching test. They contain the functionality of the test suite, each performing individual roles. Here is an example:

```

1. describe('create, edit & delete image test', () => {
2.   //...
3.
4.   it('should create image', () => {
5.     images.createImage(imageName, poolName, '1');
6.     images.getFirstTableCell(imageName).should('exist');
7.   });
8.
9.   it('should edit image', () => {
10.    images.editImage(imageName, poolName, newImageName, '2');
11.    images.getFirstTableCell(newImageName).should('exist');

```

```

12. });
13.
14. //...
15. });

```

As shown from the previous example, our `describe()` test suite is to create, edit and delete an image. Therefore, each `it()` completes one of these steps, one for creating, one for editing, and so on. Likewise, every `it()` blocks message should be in lowercase and written so long as "it" can be the prefix of the message. For example, `it('edits the test image' () => ...)` vs. `it('image edit test' () => ...)`. As shown, the first example makes grammatical sense with `it()` as the prefix whereas the second message does not. `it()` should describe what the individual test is doing and what it expects to happen.

## Differences between Frontend Unit Tests and End-to-End (E2E) Tests / FAQ

General introduction about testing and E2E/unit tests

### What are E2E/unit tests designed for?

E2E test:

It requires a fully functional system and tests the interaction of all components of the application (Ceph, back-end, front-end). E2E tests are designed to mimic the behavior of the user when interacting with the application - for example when it comes to workflows like creating/editing/deleting an item. Also the tests should verify that certain items are displayed as a user would see them when clicking through the UI (for example a menu entry or a pool that has been created during a test and the pool and its properties should be displayed in the table).

Angular Unit Tests:

Unit tests, as the name suggests, are tests for smaller units of the code. Those tests are designed for testing all kinds of Angular components (e.g. services, pipes etc.). They do not require a connection to the backend, hence those tests are independent of it. The expected data of the backend is mocked in the frontend and by using this data the functionality of the frontend can be tested without having to have real data from the backend. As previously mentioned, data is either mocked or, in a simple case, contains a static input, a function call and an expected static output. More complex examples include the state of a component (attributes of the component class), that define how the output changes according to the given input.

### Which E2E/unit tests are considered to be valid?

This is not easy to answer, but new tests that are written in the same way as already existing dashboard tests should generally be considered valid. Unit tests should focus

on the component to be tested. This is either an Angular component, directive, service, pipe, etc.

E2E tests should focus on testing the functionality of the whole application. Approximately a third of the overall E2E tests should verify the correctness of user visible elements.

## How should an E2E/unit test look like?

Unit tests should focus on the described purpose and shouldn't try to test other things in the same it block.

E2E tests should contain a description that either verifies the correctness of a user visible element or a complete process like for example the creation/validation/deletion of a pool.

## What should an E2E/unit test cover?

E2E tests should mostly, but not exclusively, cover interaction with the backend. This way the interaction with the backend is utilized to write integration tests.

A unit test should mostly cover critical or complex functionality of a component (Angular Components, Services, Pipes, Directives, etc).

## What should an E2E/unit test NOT cover?

Avoid duplicate testing: do not write E2E tests for what's already been covered as frontend-unit tests and vice versa. It may not be possible to completely avoid an overlap.

Unit tests should not be used to extensively click through components and E2E tests shouldn't be used to extensively test a single component of Angular.

## Best practices/guideline

As a general guideline we try to follow the 70/20/10 approach - 70% unit tests, 20% integration tests and 10% end-to-end tests. For further information please refer to [this document](#) and the included "Testing Pyramid".

## Further Help

To get more help on the Angular CLI use `ng help` or go check out the [Angular CLI README](#).

## Example of a Generator

```
1. # Create module 'Core'  
2. src/app> ng generate module core -m=app --routing
```

```
3.  
4. # Create module 'Auth' under module 'Core'  
5. src/app/core> ng generate module auth -m=core --routing  
6. or, alternatively:  
7. src/app> ng generate module core/auth -m=core --routing  
8.  
9. # Create component 'Login' under module 'Auth'  
10. src/app/core/auth> ng generate component login -m=core/auth  
11. or, alternatively:  
12. src/app> ng generate component core/auth/login -m=core/auth
```

# Frontend Typescript Code Style Guide Recommendations

Group the imports based on its source and separate them with a blank line.

The source groups can be either from Angular, external or internal.

Example:

```
1. import { Component } from '@angular/core';
2. import { Router } from '@angular/router';
3.
4. import { ToastrManager } from 'ngx-toastr';
5.
6. import { Credentials } from '../../../../../shared/models/credentials.model';
7. import { HostService } from './services/host.service';
```

## Frontend components

There are several components that can be reused on different pages. This components are declared on the components module:

src/pybind/mgr/dashboard/frontend/src/app/shared/components.

## Helper

This component should be used to provide additional information to the user.

Example:

```
1. <cd-helper>
2.   Some <strong>helper</strong> html text
3. </cd-helper>
```

## Terminology and wording

Instead of using the Ceph component names, the approach suggested is to use the logical/generic names (Block over RBD, Filesystem over CephFS, Object over RGW). Nevertheless, as Ceph-Dashboard cannot completely hide the Ceph internals, some Ceph-specific names might remain visible.

Regarding the wording for action labels and other textual elements (form titles, buttons, etc.), the chosen approach is to follow [these guidelines](#). As a rule of thumb, 'Create' and 'Delete' are the proper wording for most forms, instead of 'Add' and 'Remove', unless some already created item is either added or removed to/from a set of items (e.g.: 'Add permission' to a user vs. 'Create (new) permission').

In order to enforce the use of this wording, a service [ActionLabelsI18n](#) has been

created, which provides translated labels for use in UI elements.

## Frontend branding

Every vendor can customize the ‘Ceph dashboard’ to his needs. No matter if logo, HTML-Template or TypeScript, every file inside the frontend folder can be replaced.

To replace files, open `./frontend/angular.json` and scroll to the section `fileReplacements` inside the production configuration. Here you can add the files you wish to brand. We recommend to place the branded version of a file in the same directory as the original one and to add a `.brand` to the file name, right in front of the file extension. A `fileReplacement` could for example look like this:

```
1. {
2.   "replace": "src/app/core/auth/login/login.component.html",
3.   "with": "src/app/core/auth/login/login.component.brand.html"
4. }
```

To serve or build the branded user interface run:

```
$ npm run start --prod
```

or

```
$ npm run build --prod
```

Unfortunately it’s currently not possible to use multiple configurations when serving or building the UI at the same time. That means a configuration just for the branding `fileReplacements` is not an option, because you want to use the production configuration anyway (<https://github.com/angular/angular-cli/issues/10612>). Furthermore it’s also not possible to use glob expressions for `fileReplacements`. As long as the feature hasn’t been implemented, you have to add the file replacements manually to the angular.json file (<https://github.com/angular/angular-cli/issues/12354>).

Nevertheless you should stick to the suggested naming scheme because it makes it easier for you to use glob expressions once it’s supported in the future.

To change the variable defaults or add your own ones you can overwrite them in `./frontend/src/styles/vendor/_variables.scss`. Just reassign the variable you want to change, for example `$color-primary: teal;`. To overwrite or extend the default CSS, you can add your own styles in `./frontend/src/styles/vendor/_style-overrides.scss`.

## UI Style Guide

The style guide is created to document Ceph Dashboard standards and maintain consistency across the project. Its an effort to make it easier for contributors to process designing and deciding mockups and designs for Dashboard.

The development environment for Ceph Dashboard has live reloading enabled so any changes made in UI are reflected in open browser windows. Ceph Dashboard uses Bootstrap as the main third-party CSS library.

Avoid duplication of code. Be consistent with the existing UI by reusing existing SCSS declarations as much as possible.

Always check for existing code similar to what you want to write. You should always try to keep the same look-and-feel as the existing code.

## Colors

All the colors used in Ceph Dashboard UI are listed in `frontend/src/styles/defaults/_bootstrap-defaults.scss`. If using new color always define color variables in the `_bootstrap-defaults.scss` and use the variable instead of hard coded color values so that changes to the color are reflected in similar UI elements.

The main color for the Ceph Dashboard is `$primary`. The primary color is used in navigation components and as the `$border-color` for input components of form.

The secondary color is `$secondary` and is the background color for Ceph Dashboard.

## Buttons

Buttons are used for performing actions such as: "Submit", "Edit", "Create" and "Update".

**Forms:** When using to submit forms anywhere in the Dashboard, the main action button should use the `cd-submit-button` component and the secondary button should use `cd-back-button` component. The text on the action button should be same as the form title and follow a title case. The text on the secondary button should be Cancel. Perform action button should always be on right while Cancel button should always be on left.

**Modals:** The main action button should use the `cd-submit-button` component and the secondary button should use `cd-back-button` component. The text on the action button should follow a title case and correspond to the action to be performed. The text on the secondary button should be Close.

**Disclosure Button:** Disclosure buttons should be used to allow users to display and hide additional content in the interface.

**Action Button:** Use the action button to perform actions such as edit or update a component. All action button should have an icon corresponding to the actions they perform and button text should follow title case. The button color should be the same as the form's main button color.

**Drop Down Buttons:** Use dropdown buttons to display predefined lists of actions. All drop down buttons have icons corresponding to the action they perform.

## Links

Use text hyperlinks as navigation to guide users to a new page in the application or to anchor users to a section within a page. The color of the hyperlinks should be \$primary.

## Forms

Mark invalid form fields with red outline and show a meaningful error message. Use red as font color for message and be as specific as possible. This field is required. should be the exact error message for required fields. Mark valid forms with a green outline and a green tick at the end of the form. Sections should not have a bigger header than the parent.

## Modals

Blur any interface elements in the background to bring the modal content into focus. The heading of the modal should reflect the action it can perform and should be clearly mentioned at the top of the modal. Use cd-back-button component in the footer for closing the modal.

## Icons

We use [Fork Awesome](#) classes for icons. We have a list of used icons in src/app/shared/enum/icons.enum.ts, these should be referenced in the HTML, so its easier to change them later. When icons are next to text, they should be center-aligned horizontally. If icons are stacked, they should also be center-aligned vertically. Use small icons with buttons. For notifications use large icons.

## Navigation

For local navigation use tabs. For overall navigation use expandable vertical navigation to collapse and expand items as needed.

## Alerts and notifications

Default notification should have text-info color. Success notification should have text-success color. Failure notification should have text-danger color.

## I18N

---

## How to extract messages from source code?

To extract the I18N messages from the templates and the TypeScript files just run the following command in `src/pybind/mgr/dashboard/frontend` :

```
1. $ npm run i18n:extract
```

This will extract all marked messages from the HTML templates first and then add all marked strings from the TypeScript files to the translation template. Since the extraction from TypeScript files is still not supported by Angular itself, we are using the [ngx-translator](#) extractor to parse the TypeScript files.

When the command ran successfully, it should have created or updated the file

```
src/locale/messages.xlf .
```

The file isn't tracked by git, you can just use it to start with the translation offline or add/update the resource files on [transifex](#).

## Supported languages

All our supported languages should be registered in both exports in [supported-languages.enum.ts](#) and have a corresponding test in [language-selector.component.spec.ts](#).

The [SupportedLanguages](#) enum will provide the list for the default language selection.

## Translating process

To facilitate the translation process of the dashboard we are using a web tool called [transifex](#).

If you wish to help translating to any language just go to our [transifex project page](#), join the project and you can start translating immediately.

All translations will then be reviewed and later pushed upstream.

## Updating translated messages

Any time there are new messages translated and reviewed in a specific language we should update the translation file upstream.

To do that, check the settings in the i18n config file

```
src/pybind/mgr/dashboard/frontend/i18n.config.json ::
```

and make sure that the organization is *ceph*, the project is *ceph-dashboard* and the resource is the one you want to pull from and push to e.g. *Master:master*. To find a list of available resources visit <https://www.transifex.com/ceph/ceph-dashboard/content/>.

After you checked the config go to the directory [src/pybind/mgr/dashboard/frontend](#) and run:

```
1. $ npm run i18n
```

This command will extract all marked messages from the HTML templates and TypeScript files. Once the source file has been created it will push it to transifex and pull the

latest translations. It will also fill all the untranslated strings with the source string. The tool will ask you for an api token, unless you added it by running:

```
$ npm run i18n:token
```

To create a transifex api token visit <https://www.transifex.com/user/settings/api/>.

After the command ran successfully, build the UI and check if everything is working as expected. You also might want to run the frontend tests.

## Suggestions

Strings need to start and end in the same line as the element:

```
1. <!-- avoid -->
2. <span i18n>
3.   Foo
4. </span>
5.
6. <!-- recommended -->
7. <span i18n>Foo</span>
8.
9.
10. <!-- avoid -->
11. <span i18n>
12.   Foo bar baz.
13.   Foo bar baz.
14. </span>
15.
16. <!-- recommended -->
17. <span i18n>Foo bar baz.
18.   Foo bar baz.</span>
```

Isolated interpolations should not be translated:

```
1. <!-- avoid -->
2. <span i18n>{{ foo }}</span>
3.
4. <!-- recommended -->
5. <span>{{ foo }}</span>
```

Interpolations used in a sentence should be kept in the translation:

```
1. <!-- recommended -->
2. <span i18n>There are {{ x }} OSDs.</span>
```

Remove elements that are outside the context of the translation:

```
1. <!-- avoid -->
```

```

2. <label i18n>
3.   Profile
4.   <span class="required"></span>
5. </label>
6.
7. <!-- recommended -->
8. <label>
9.   <ng-container i18n>Profile<ng-container>
10.  <span class="required"></span>
11. </label>

```

Keep elements that affect the sentence:

```

1. <!-- recommended -->
2. <span i18n>Profile <b>foo</b> will be removed.</span>

```

## Backend Development

The Python backend code of this module requires a number of Python modules to be installed. They are listed in file `requirements.txt`. Using `pip` you may install all required dependencies by issuing `pip install -r requirements.txt` in directory `src/pybind/mgr/dashboard`.

If you're using the [ceph-dev-docker development environment](#), simply run `./install_deps.sh` from the toplevel directory to install them.

## Unit Testing

In dashboard we have two different kinds of backend tests:

1. Unit tests based on `tox`
2. API tests based on Teuthology.

### Unit tests based on tox

We included a `tox` configuration file that will run the unit tests under Python 2 or 3, as well as linting tools to guarantee the uniformity of code.

You need to install `tox` and `coverage` before running it. To install the packages in your system, either install it via your operating system's package management tools, e.g. by running `dnf install python-tox python-coverage` on Fedora Linux.

Alternatively, you can use Python's native package installation method:

```

1. $ pip install tox
2. $ pip install coverage

```

To run the tests, run `src/script/run_tox.sh` in the dashboard directory (where `tox.ini` is located):

```
1. ## Run Python 2+3 tests+lint commands:
2. $ ../../script/run_tox.sh --tox-env py27,py3 lint,check
3.
4. ## Run Python 3 tests+lint commands:
5. $ ../../script/run_tox.sh --tox-env py3 lint,check
6.
7. ## Run Python 3 arbitrary command (e.g. 1 single test):
8. $ ../../script/run_tox.sh --tox-env py3 "" tests/test_rgw_client.py::RgwClientTest::test_ssl_verify
```

You can also run tox instead of `run_tox.sh` :

```
1. ## Run Python 3 tests command:
2. $ tox -e py3
3.
4. ## Run Python 3 arbitrary command (e.g. 1 single test):
5. $ tox -e py3 tests/test_rgw_client.py::RgwClientTest::test_ssl_verify
```

Python files can be automatically fixed and formatted according to PEP8 standards by using `run_tox.sh --tox-env fix` or `tox -e fix .`

We also collect coverage information from the backend code when you run tests. You can check the coverage information provided by the tox output, or by running the following command after tox has finished successfully:

```
1. $ coverage html
```

This command will create a directory `htmlcov` with an HTML representation of the code coverage of the backend.

## API tests based on Teuthology

How to run existing API tests:

To run the API tests against a real Ceph cluster, we leverage the Teuthology framework. This has the advantage of catching bugs originated from changes in the internal Ceph code.

Our `run-backend-api-tests.sh` script will start a `vstart` Ceph cluster before running the Teuthology tests, and then it stops the cluster after the tests are run. Of course this implies that you have built/compiled Ceph previously.

Start all dashboard tests by running:

```
1. $ ./run-backend-api-tests.sh
```

Or, start one or multiple specific tests by specifying the test name:

```
1. $ ./run-backend-api-tests.sh tasks.mgr.dashboard.test_pool.PoolTest
```

Or, `source` the script and run the tests manually:

```
1. $ source run-backend-api-tests.sh
2. $ run_teuthology_tests [tests]...
3. $ cleanup_teuthology
```

How to write your own tests:

There are two possible ways to write your own API tests:

The first is by extending one of the existing test classes in the `qa/tasks/mgr/dashboard` directory.

The second way is by adding your own API test module if you're creating a new controller for example. To do so you'll just need to add the file containing your new test class to the `qa/tasks/mgr/dashboard` directory and implement all your tests here.

#### Note

Don't forget to add the path of the newly created module to `modules` section in `qa/suites/rados/mgr/tasks/dashboard.yaml`.

Short example: Let's assume you created a new controller called `my_new_controller.py` and the related test module `test_my_new_controller.py`. You'll need to add `tasks.mgr.dashboard.test_my_new_controller` to the `modules` section in the `dashboard.yaml` file.

Also, if you're removing test modules please keep in mind to remove the related section. Otherwise the Teuthology test run will fail.

Please run your API tests on your dev environment (as explained above) before submitting a pull request. Also make sure that a full QA run in Teuthology/sepiabot (based on your changes) has completed successfully before it gets merged. You don't need to schedule the QA run yourself, just add the 'needs-qa' label to your pull request as soon as you think it's ready for merging (e.g. make check was successful, the pull request is approved and all comments have been addressed). One of the developers who has access to Teuthology/the sepiabot will take care of it and report the result back to you.

## How to add a new controller?

A controller is a Python class that extends from the  `BaseController` class and is decorated with either the `@Controller`, `@ApiController` or `@UiApiController` decorators. The Python class must be stored inside a Python file located under the `controllers` directory. The Dashboard module will automatically load your new controller upon

start.

`@ApiController` and `@UiApiController` are both specializations of the `@Controller` decorator.

The `@ApiController` should be used for controllers that provide an API-like REST interface and the `@UiApiController` should be used for endpoints consumed by the UI but that are not part of the ‘public’ API. For any other kinds of controllers the `@Controller` decorator should be used.

A controller has a URL prefix path associated that is specified in the controller decorator, and all endpoints exposed by the controller will share the same URL prefix path.

A controller’s endpoint is exposed by implementing a method on the controller class decorated with the `@Endpoint` decorator.

For example create a file `ping.py` under `controllers` directory with the following code:

```

1. from ..tools import Controller, ApiController, UiApiController, BaseController, Endpoint
2.
3. @Controller('/ping')
4. class Ping(BaseController):
5.     @Endpoint()
6.     def hello(self):
7.         return {'msg': "Hello"}
8.
9. @ApiController('/ping')
10. class ApiPing(BaseController):
11.     @Endpoint()
12.     def hello(self):
13.         return {'msg': "Hello"}
14.
15. @UiApiController('/ping')
16. class UiApiPing(BaseController):
17.     @Endpoint()
18.     def hello(self):
19.         return {'msg': "Hello"}
```

The `hello` endpoint of the `Ping` controller can be reached by the following URL: [https://mgr\\_hostname:8443/ping/hello](https://mgr_hostname:8443/ping/hello) using HTTP GET requests. As you can see the controller URL path `/ping` is concatenated to the method name `hello` to generate the endpoint’s URL.

In the case of the `ApiPing` controller, the `hello` endpoint can be reached by the following URL: [https://mgr\\_hostname:8443/api/ping/hello](https://mgr_hostname:8443/api/ping/hello) using a HTTP GET request. The API controller URL path `/ping` is prefixed by the `/api` path and then concatenated to the method name `hello` to generate the endpoint’s URL. Internally, the `@ApiController` is actually calling the `@Controller` decorator by passing an additional decorator

parameter called `base_url` :

```
1. @ApiController('/ping') <=> @Controller('/ping', base_url="/api")
```

`UiApiPing` works in a similar way than the `ApiPing`, but the URL will be prefixed by `/ui-api` : `https://mgr_hostname:8443/ui-api/ping/hello`. `UiApiPing` is also a `@Controller` extension:

```
1. @UiApiController('/ping') <=> @Controller('/ping', base_url="/ui-api")
```

The `@Endpoint` decorator also supports many parameters to customize the endpoint:

- `method="GET"` : the HTTP method allowed to access this endpoint.
- `path="/<method_name>"` : the URL path of the endpoint, excluding the controller URL path prefix.
- `path_params=[]` : list of method parameter names that correspond to URL path parameters. Can only be used when `method in ['POST', 'PUT']`.
- `query_params=[]` : list of method parameter names that correspond to URL query parameters.
- `json_response=True` : indicates if the endpoint response should be serialized in JSON format.
- `proxy=False` : indicates if the endpoint should be used as a proxy.

An endpoint method may have parameters declared. Depending on the HTTP method defined for the endpoint the method parameters might be considered either path parameters, query parameters, or body parameters.

For `GET` and `DELETE` methods, the method's non-optional parameters are considered path parameters by default. Optional parameters are considered query parameters. By specifying the `query_parameters` in the endpoint decorator it is possible to make a non-optional parameter to be a query parameter.

For `POST` and `PUT` methods, all method parameters are considered body parameters by default. To override this default, one can use the `path_params` and `query_params` to specify which method parameters are path and query parameters respectively. Body parameters are decoded from the request body, either from a form format, or from a dictionary in JSON format.

Let's use an example to better understand the possible ways to customize an endpoint:

```
1. from ..tools import Controller, BaseController, Endpoint
2.
3. @Controller('/ping')
4. class Ping(BaseController):
```

```

5.
6. # URL: /ping/{key}?opt1=...&opt2=...
7. @Endpoint(path="/", query_params=['opt1'])
8. def index(self, key, opt1, opt2=None):
9.     """
10.
11. # URL: /ping/{key}?opt1=...&opt2=...
12. @Endpoint(query_params=['opt1'])
13. def __call__(self, key, opt1, opt2=None):
14.     """
15.
16. # URL: /ping/post/{key1}/{key2}
17. @Endpoint('POST', path_params=['key1', 'key2'])
18. def post(self, key1, key2, data1, data2=None):
19.     """

```

In the above example we see how the `path` option can be used to override the generated endpoint URL in order to not use the method's name in the URL. In the `index` method we set the `path` to `"/"` to generate an endpoint that is accessible by the root URL of the controller.

An alternative approach to generate an endpoint that is accessible through just the controller's path URL is by using the `__call__` method, as we show in the above example.

From the third method you can see that the path parameters are collected from the URL by parsing the list of values separated by slashes `/` that come after the URL path `/ping` for `index` method case, and `/ping/post` for the `post` method case.

Defining path parameters in endpoints's URLs using python methods's parameters is very easy but it is still a bit strict with respect to the position of these parameters in the URL structure. Sometimes we may want to explicitly define a URL scheme that contains path parameters mixed with static parts of the URL. Our controller infrastructure also supports the declaration of URL paths with explicit path parameters at both the controller level and method level.

Consider the following example:

```

1. from ..tools import Controller, BaseController, Endpoint
2.
3. @Controller('/ping/{node}/stats')
4. class Ping(BaseController):
5.
6.     # URL: /ping/{node}/stats/{date}/latency?unit=...
7.     @Endpoint(path="/{date}/latency")
8.     def latency(self, node, date, unit="ms"):
9.         """

```

In this example we explicitly declare a path parameter `{node}` in the controller URL path, and a path parameter `{date}` in the `latency` method. The endpoint for the

`latency` method is then accessible through the URL:  
`https://mgr_hostname:8443/ping/{node}/stats/{date}/latency` .

For a full set of examples on how to use the `@Endpoint` decorator please check the unit test file: `tests/test_controllers.py` . There you will find many examples of how to customize endpoint methods.

## Implementing Proxy Controller

Sometimes you might need to relay some requests from the Dashboard frontend directly to an external service. For that purpose we provide a decorator called `@Proxy` . (As a concrete example, check the `controllers/rgw.py` file where we implemented an RGW Admin Ops proxy.)

The `@Proxy` decorator is a wrapper of the `@Endpoint` decorator that already customizes the endpoint for working as a proxy. A proxy endpoint works by capturing the URL path that follows the controller URL prefix path, and does not do any decoding of the request body.

Example:

```

1. from ..tools import Controller, BaseController, Proxy
2.
3. @Controller('/foo/proxy')
4. class FooServiceProxy(BaseController):
5.
6.     @Proxy()
7.     def proxy(self, path, **params):
8.         """
9.         if requested URL is "/foo/proxy/access/service?opt=1"
10.            then path is "access/service" and params is {'opt': '1'}
11.        """

```

## How does the RESTController work?

We also provide a simple mechanism to create REST based controllers using the `RESTController` class. Any class which inherits from `RESTController` will, by default, return JSON.

The `RESTController` is basically an additional abstraction layer which eases and unifies the work with collections. A collection is just an array of objects with a specific type. `RESTController` enables some default mappings of request types and given parameters to specific method names. This may sound complicated at first, but it's fairly easy. Lets have look at the following example:

```

1. import cherrypy
2. from ..tools import ApiController, RESTController
3.
```

```

4. @ApiController('ping')
5. class Ping(RESTController):
6.     def list(self):
7.         return {"msg": "Hello"}
8.
9.     def get(self, id):
10.        return self.objects[id]

```

In this case, the `list` method is automatically used for all requests to `api/ping` where no additional argument is given and where the request type is `GET`. If the request is given an additional argument, the ID in our case, it won't map to `list` anymore but to `get` and return the element with the given ID (assuming that `self.objects` has been filled before). The same applies to other request types:

Request type	Arguments	Method	Status Code
GET	No	list	200
PUT	No	bulk_set	200
POST	No	create	201
DELETE	No	bulk_delete	204
GET	Yes	get	200
PUT	Yes	set	200
DELETE	Yes	delete	204

## How to use a custom API endpoint in a RESTController?

If you don't have any access restriction you can use `@Endpoint`. If you have set a permission scope to restrict access to your endpoints, `@Endpoint` will fail, as it doesn't know which permission property should be used. To use a custom endpoint inside a restricted `RESTController` use `@RESTController.Collection` instead. You can also choose `@RESTController.Resource` if you have set a `RESOURCE_ID` in your `RESTController` class.

```

1. import cherrypy
2. from ..tools import ApiController, RESTController
3.
4. @ApiController('ping', Scope.Ping)
5. class Ping(RESTController):
6.     RESOURCE_ID = 'ping'
7.
8.     @RESTController.Resource('GET')

```

```

9.     def some_get_endpoint(self):
10.        return {"msg": "Hello"}
11.
12.    @RESTController.Collection('POST')
13.    def some_post_endpoint(self, **data):
14.        return {"msg": data}

```

Both decorators also support four parameters to customize the endpoint:

- `method="GET"` : the HTTP method allowed to access this endpoint.
- `path="/<method_name>"` : the URL path of the endpoint, excluding the controller URL path prefix.
- `status=200` : set the HTTP status response code
- `query_params=[]` : list of method parameter names that correspond to URL query parameters.

## How to restrict access to a controller?

All controllers require authentication by default. If you require that the controller can be accessed without authentication, then you can add the parameter `secure=False` to the controller decorator.

Example:

```

1. import cherrypy
2. from . import ApiController, RESTController
3.
4.
5. @ApiController('ping', secure=False)
6. class Ping(RESTController):
7.     def list(self):
8.         return {"msg": "Hello"}

```

## How to create a dedicated UI endpoint which uses the 'public' API?

Sometimes we want to combine multiple calls into one single call to save bandwidth or for other performance reasons. In order to achieve that, we first have to create an `@UiApiController` which is used for endpoints consumed by the UI but that are not part of the 'public' API. Let the ui class inherit from the REST controller class. Now you can use all methods from the api controller.

Example:

```

1. import cherrypy
2. from . import UiApiController, ApiController, RESTController

```

```

3.
4.
5. @ApiController('ping', secure=False) # /api/ping
6. class Ping(RESTController):
7.     def list(self):
8.         return self._list()
9.
10.    def _list(self): # To not get in conflict with the JSON wrapper
11.        return [1, 2, 3]
12.
13.
14. @UiApiController('ping', secure=False) # /ui-api/ping
15. class PingUi(Ping):
16.     def list(self):
17.         return self._list() + [4, 5, 6]

```

## How to access the manager module instance from a controller?

We provide the manager module instance as a global variable that can be imported in any module.

Example:

```

1. import logging
2. import cherrypy
3. from .. import mgr
4. from ..tools import ApiController, RESTController
5.
6. logger = logging.getLogger(__name__)
7.
8. @ApiController('servers')
9. class Servers(RESTController):
10.    def list(self):
11.        logger.debug('Listing available servers')
12.        return {'servers': mgr.list_servers()}

```

## How to write a unit test for a controller?

We provide a test helper class called `ControllerTestCase` to easily create unit tests for your controller.

If we want to write a unit test for the above `Ping` controller, create a `test_ping.py` file under the `tests` directory with the following code:

```

1. from .helper import ControllerTestCase
2. from .controllers.ping import Ping
3.
4.

```

```

5. class PingTest(ControllerTestCase):
6.     @classmethod
7.     def setup_test(cls):
8.         Ping._cp_config['tools.authenticate.on'] = False
9.         cls.setup_controllers([Ping])
10.
11.    def test_ping(self):
12.        self._get("/api/ping")
13.        self.assertStatus(200)
14.        self.assertJsonBody({'msg': 'Hello'})

```

The `ControllerTestCase` class starts by initializing a CherryPy webserver. Then it will call the `setup_test()` class method where we can explicitly load the controllers that we want to test. In the above example we are only loading the `Ping` controller. We can also disable authentication of a controller at this stage, as depicted in the example.

## How to listen for manager notifications in a controller?

The manager notifies the modules of several types of cluster events, such as cluster logging event, etc...

Each module has a “global” handler function called `notify` that the manager calls to notify the module. But this handler function must not block or spend too much time processing the event notification. For this reason we provide a notification queue that controllers can register themselves with to receive cluster notifications.

The example below represents a controller that implements a very simple live log viewer page:

```

1. from __future__ import absolute_import
2.
3. import collections
4.
5. import cherrypy
6.
7. from ..tools import ApiController, BaseController, NotificationQueue
8.
9.
10. @ApiController('livelog')
11. class LiveLog(BaseController):
12.     log_buffer = collections.deque(maxlen=1000)
13.
14.     def __init__(self):
15.         super(LiveLog, self).__init__()
16.         NotificationQueue.register(self.log, 'clog')
17.
18.     def log(self, log_struct):
19.         self.log_buffer.appendleft(log_struct)
20.

```

```

21.     @cherrypy.expose
22.     def default(self):
23.         ret = '<html><meta http-equiv="refresh" content="2" /><body>'
24.         for l in self.log_buffer:
25.             ret += "{}<br>".format(l)
26.         ret += "</body></html>"
27.     return ret

```

As you can see above, the `NotificationQueue` class provides a register method that receives the function as its first argument, and receives the “notification type” as the second argument. You can omit the second argument of the `register` method, and in that case you are registering to listen all notifications of any type.

Here is an list of notification types (these might change in the future) that can be used:

- `clog` : cluster log notifications
- `command` : notification when a command issued by `MgrModule.send_command` completes
- `perf_schema_update` : perf counters schema update
- `mon_map` : monitor map update
- `fs_map` : cephfs map update
- `osd_map` : OSD map update
- `service_map` : services (RGW, RBD-Mirror, etc.) map update
- `mon_status` : monitor status regular update
- `health` : health status regular update
- `pg_summary` : regular update of PG status information

## How to write a unit test when a controller accesses a Ceph module?

Consider the following example that implements a controller that retrieves the list of RBD images of the `rbd` pool:

```

1. import rbd
2. from .. import mgr
3. from ..tools import ApiController, RESTController
4.
5.
6. @ApiController('rbdimages')
7. class RbdImages(RESTController):
8.     def __init__(self):
9.         self.ioctx = mgr.rados.open_ioctx('rbd')

```

```

10.     self.rbd = rbd.RBD()
11.
12.     def list(self):
13.         return [{name: n} for n in self.rbd.list(self.ioctx)]

```

In the example above, we want to mock the return value of the `rbd.list` function, so that we can test the JSON response of the controller.

The unit test code will look like the following:

```

1. import mock
2. from .helper import ControllerTestCase
3.
4.
5. class RbdImagesTest(ControllerTestCase):
6.     @mock.patch('rbd.RBD.list')
7.     def test_list(self, rbd_list_mock):
8.         rbd_list_mock.return_value = ['img1', 'img2']
9.         self._get('/api/rbdimages')
10.        self.assertJsonBody([{'name': 'img1'}, {'name': 'img2'}])

```

## How to add a new configuration setting?

If you need to store some configuration setting for a new feature, we already provide an easy mechanism for you to specify/use the new config setting.

For instance, if you want to add a new configuration setting to hold the email address of the dashboard admin, just add a setting name as a class attribute to the `Options` class in the `settings.py` file:

```

1. # ...
2. class Options(object):
3.     # ...
4.
5. ADMIN_EMAIL_ADDRESS = ('admin@admin.com', str)

```

The value of the class attribute is a pair composed by the default value for that setting, and the python type of the value.

By declaring the `ADMIN_EMAIL_ADDRESS` class attribute, when you restart the dashboard module, you will automatically gain two additional CLI commands to get and set that setting:

```

1. $ ceph dashboard get-admin-email-address
2. $ ceph dashboard set-admin-email-address <value>

```

To access, or modify the config setting value from your Python code, either inside a controller or anywhere else, you just need to import the `Settings` class and access it

like this:

```

1. from settings import Settings
2.
3. # ...
4. tmp_var = Settings.ADMIN_EMAIL_ADDRESS
5.
6. # ....
7. Settings.ADMIN_EMAIL_ADDRESS = 'myemail@admin.com'
```

The settings management implementation will make sure that if you change a setting value from the Python code you will see that change when accessing that setting from the CLI and vice-versa.

## How to run a controller read-write operation asynchronously?

Some controllers might need to execute operations that alter the state of the Ceph cluster. These operations might take some time to execute and to maintain a good user experience in the Web UI, we need to run those operations asynchronously and return immediately to frontend some information that the operations are running in the background.

To help in the development of the above scenario we added the support for asynchronous tasks. To trigger the execution of an asynchronous task we must use the following class method of the `TaskManager` class:

```

1. from ..tools import TaskManager
2. # ...
3. TaskManager.run(name, metadata, func, args, kwargs)
```

- `name` is a string that can be used to group tasks. For instance for RBD image creation tasks we could specify `"rbd/create"` as the name, or similarly `"rbd/remove"` for RBD image removal tasks.
- `metadata` is a dictionary where we can store key-value pairs that characterize the task. For instance, when creating a task for creating RBD images we can specify the metadata argument as `{'pool_name': "rbd", image_name': "test-img"}`.
- `func` is the python function that implements the operation code, which will be executed asynchronously.
- `args` and `kwargs` are the positional and named arguments that will be passed to `func` when the task manager starts its execution.

The `TaskManager.run` method triggers the asynchronous execution of function `func` and returns a `Task` object. The `Task` provides the public method `Task.wait(timeout)`, which can be used to wait for the task to complete up to a timeout defined in seconds and

provided as an argument. If no argument is provided the `wait` method blocks until the task is finished.

The `Task.wait` is very useful for tasks that usually are fast to execute but that sometimes may take a long time to run. The return value of the `Task.wait` method is a pair `(state, value)` where `state` is a string with following possible values:

- `VALUE_DONE = "done"`
- `VALUE_EXECUTING = "executing"`

The `value` will store the result of the execution of function `func` if `state == VALUE_DONE`. If `state == VALUE_EXECUTING` then `value == None`.

The pair `(name, metadata)` should unequivocally identify the task being run, which means that if you try to trigger a new task that matches the same `(name, metadata)` pair of the currently running task, then the new task is not created and you get the task object of the current running task.

For instance, consider the following example:

```
1. task1 = TaskManager.run("dummy/task", {'attr': 2}, func)
2. task2 = TaskManager.run("dummy/task", {'attr': 2}, func)
```

If the second call to `TaskManager.run` executes while the first task is still executing then it will return the same task object: `assert task1 == task2`.

## How to get the list of executing and finished asynchronous tasks?

The list of executing and finished tasks is included in the `Summary` controller, which is already polled every 5 seconds by the dashboard frontend. But we also provide a dedicated controller to get the same list of executing and finished tasks.

The `Task` controller exposes the `/api/task` endpoint that returns the list of executing and finished tasks. This endpoint accepts the `name` parameter that accepts a glob expression as its value. For instance, an HTTP GET request of the URL `/api/task?name=rbd/*` will return all executing and finished tasks which name starts with `rbd`.

To prevent the finished tasks list from growing unbounded, we will always maintain the 10 most recent finished tasks, and the remaining older finished tasks will be removed when reaching a TTL of 1 minute. The TTL is calculated using the timestamp when the task finished its execution. After a minute, when the finished task information is retrieved, either by the summary controller or by the task controller, it is automatically deleted from the list and it will not be included in further task queries.

Each executing task is represented by the following dictionary:

```

1. {
2.     'name': "name",    # str
3.     'metadata': { },   # dict
4.     'begin_time': "2018-03-14T15:31:38.423605Z",  # str (ISO 8601 format)
5.     'progress': 0    # int (percentage)
6. }
```

Each finished task is represented by the following dictionary:

```

1. {
2.     'name': "name",    # str
3.     'metadata': { },   # dict
4.     'begin_time': "2018-03-14T15:31:38.423605Z",  # str (ISO 8601 format)
5.     'end_time': "2018-03-14T15:31:39.423605Z",  # str (ISO 8601 format)
6.     'duration': 0.0,  # float
7.     'progress': 0    # int (percentage)
8.     'success': True,  # bool
9.     'ret_value': None, # object, populated only if 'success' == True
10.    'exception': None, # str, populated only if 'success' == False
11. }
```

## How to use asynchronous APIs with asynchronous tasks?

The `TaskManager.run` method as described in a previous section, is well suited for calling blocking functions, as it runs the function inside a newly created thread. But sometimes we want to call some function of an API that is already asynchronous by nature.

For these cases we want to avoid creating a new thread for just running a non-blocking function, and want to leverage the asynchronous nature of the function. The

`TaskManager.run` is already prepared to be used with non-blocking functions by passing an object of the type `TaskExecutor` as an additional parameter called `executor`. The full method signature of `TaskManager.run` :

```
1. TaskManager.run(name, metadata, func, args=None, kwargs=None, executor=None)
```

The `TaskExecutor` class is responsible for code that executes a given task function, and defines three methods that can be overridden by subclasses:

```

1. def init(self, task)
2. def start(self)
3. def finish(self, ret_value, exception)
```

The `init` method is called before the running the task function, and receives the task object (of class `Task` ).

The `start` method runs the task function. The default implementation is to run the task function in the current thread context.

The `finish` method should be called when the task function finishes with either the `ret_value` populated with the result of the execution, or with an exception object in the case that execution raised an exception.

To leverage the asynchronous nature of a non-blocking function, the developer should implement a custom executor by creating a subclass of the `TaskExecutor` class, and provide an instance of the custom executor class as the `executor` parameter of the `TaskManager.run`.

To better understand the expressive power of executors, we write a full example of use a custom executor to execute the `MgrModule.send_command` asynchronous function:

```

1. import json
2. from mgr_module import CommandResult
3. from .. import mgr
4. from ..tools import ApiController, RESTController, NotificationQueue, \
5.                     TaskManager, TaskExecutor
6.
7.
8. class SendCommandExecutor(TaskExecutor):
9.     def __init__(self):
10.         super(SendCommandExecutor, self).__init__()
11.         self.tag = None
12.         self.result = None
13.
14.     def init(self, task):
15.         super(SendCommandExecutor, self).init(task)
16.
17.         # we need to listen for 'command' events to know when the command
18.         # finishes
19.         NotificationQueue.register(self._handler, 'command')
20.
21.         # store the CommandResult object to retrieve the results
22.         self.result = self.task.fn_args[0]
23.         if len(self.task.fn_args) > 4:
24.             # the user specified a tag for the command, so let's use it
25.             self.tag = self.task.fn_args[4]
26.         else:
27.             # let's generate a unique tag for the command
28.             self.tag = 'send_command_{}'.format(id(self))
29.             self.task.fn_args.append(self.tag)
30.
31.     def _handler(self, data):
32.         if data == self.tag:
33.             # the command has finished, notifying the task with the result
34.             self.finish(self.result.wait(), None)
35.             # deregister listener to avoid memory leaks
36.             NotificationQueue.deregister(self._handler, 'command')
37.

```

```

38.
39. @ApiController('test')
40. class Test(RESTController):
41.
42.     def _run_task(self, osd_id):
43.         task = TaskManager.run("test/task", {}, mgr.send_command,
44.                               [CommandResult(''), 'osd', osd_id,
45.                                json.dumps({'prefix': 'perf histogram dump'})]),
46.                               executor=SendCommandExecutor())
47.         return task.wait(1.0)
48.
49.     def get(self, osd_id):
50.         status, value = self._run_task(osd_id)
51.         return {'status': status, 'value': value}

```

The above `SendCommandExecutor` executor class can be used for any call to `MgrModule.send_command`. This means that we should need just one custom executor class implementation for each non-blocking API that we use in our controllers.

The default executor, used when no executor object is passed to `TaskManager.run`, is the `ThreadedExecutor`. You can check its implementation in the `tools.py` file.

## How to update the execution progress of an asynchronous task?

The asynchronous tasks infrastructure provides support for updating the execution progress of an executing task. The progress can be updated from within the code the task is executing, which usually is the place where we have the progress information available.

To update the progress from within the task code, the `TaskManager` class provides a method to retrieve the current task object:

```
1. TaskManager.current_task()
```

The above method is only available when using the default executor `ThreadedExecutor` for executing the task. The `current_task()` method returns the current `Task` object. The `Task` object provides two public methods to update the execution progress value: the `set_progress(percentage)`, and the `inc_progress(delta)` methods.

The `set_progress` method receives as argument an integer value representing the absolute percentage that we want to set to the task.

The `inc_progress` method receives as argument an integer value representing the delta we want to increment to the current execution progress percentage.

Take the following example of a controller that triggers a new task and updates its progress:

```

1. from __future__ import absolute_import
2. import random
3. import time
4. import cherrypy
5. from ..tools import TaskManager, ApiController, BaseController
6.
7.
8. @ApiController('dummy_task')
9. class DummyTask(BaseController):
10.     def _dummy(self):
11.         top = random.randrange(100)
12.         for i in range(top):
13.             TaskManager.current_task().set_progress(i*100/top)
14.             # or TaskManager.current_task().inc_progress(100/top)
15.             time.sleep(1)
16.         return "finished"
17.
18.     @cherrypy.expose
19.     @cherrypy.tools.json_out()
20.     def default(self):
21.         task = TaskManager.run("dummy/task", {}, self._dummy)
22.         return task.wait(5) # wait for five seconds

```

## How to deal with asynchronous tasks in the front-end?

All executing and most recently finished asynchronous tasks are displayed on “Background-Tasks” and if finished on “Recent-Notifications” in the menu bar. For each task a operation name for three states (running, success and failure), a function that tells who is involved and error descriptions, if any, have to be provided. This can be achieved by appending `TaskManagerMessageService.messages`. This has to be done to achieve consistency among all tasks and states.

### Operation Object

Ensures consistency among all tasks. It consists of three verbs for each different state f.e. `{running: 'Creating', failure: 'create', success: 'Created'}`.

1. Put running operations in present participle f.e. `'Updating'`.
2. Failed messages always start with `'Failed to '` and should be continued with the operation in present tense f.e. `'update'`.
3. Put successful operations in past tense f.e. `'Updated'`.

### Involves Function

Ensures consistency among all messages of a task, it resembles who's involved by the operation. It's a function that returns a string which takes the metadata from the task to return f.e. `"RBD 'somePool/someImage'"`.

Both combined create the following messages:

- Failure => "Failed to create RBD 'somePool/someImage'"
- Running => "Creating RBD 'somePool/someImage'"
- Success => "Created RBD 'somePool/someImage'"

For automatic task handling use `TaskWrapperService.wrapTaskAroundCall` .

If for some reason `wrapTaskAroundCall` is not working for you, you have to subscribe to your asynchronous task manually through `TaskManagerService.subscribe` , and provide it with a callback, in case of a success to notify the user. A notification can be triggered with `NotificationService.notifyTask` . It will use `TaskManagerMessageService.messages` to display a message based on the state of a task.

Notifications of API errors are handled by `ApiInterceptorService` .

Usage example:

```

1. export class TaskManagerMessageService {
2.   // ...
3.   messages = {
4.     // Messages for task 'rbd/create'
5.     'rbd/create': new TaskManagerMessage(
6.       // Message prefixes
7.       ['create', 'Creating', 'Created'],
8.       // Message suffix
9.       (metadata) => `RBD '${metadata.pool_name}/${metadata.image_name}'`,
10.      (metadata) => ({
11.        // Error code and description
12.        '17': `Name is already used by RBD '${metadata.pool_name}/${{
13.          metadata.image_name}'` .
14.      })
15.    ),
16.    // ...
17.  };
18.  // ...
19. }
20.
21. export class RBDFormComponent {
22.   // ...
23.   createAction() {
24.     const request = this.createRequest();
25.     // Subscribes to 'call' with submitted 'task' and handles notifications
26.     return this.taskWrapper.wrapTaskAroundCall({
27.       task: new FinishedTask('rbd/create', {
28.         pool_name: request.pool_name,
29.         image_name: request.name
30.       }),
31.       call: this.rbdService.create(request)
32.     });

```

```

33.    }
34.    // ...
35. }
```

## REST API documentation

Ceph-Dashboard provides two types of documentation for the **Ceph RESTful API**:

- **Static documentation:** available at [Ceph RESTful API](#). This comes from a versioned specification located at `src/pybind/mgr/dashboard/openapi.yaml` .
- **Interactive documentation:** available from a running Ceph-Dashboard instance (top-right `?` icon > API Docs).

If changes are made to the `controllers/` directory, it's very likely that they will result in changes to the generated OpenAPI specification. For that reason, a checker has been implemented to block unintended changes. This check is automatically triggered by the Pull Request CI (`make check`) and can be also manually invoked: `tox -e openapi-check` .

If that checker failed, it means that the current Pull Request is modifying the Ceph API and therefore:

1. The versioned OpenAPI specification should be updated explicitly: `tox -e openapi-fix` .
2. The team @ceph/api will be requested for reviews (this is automated via Github CODEOWNERS), in order to asses the impact of changes.

Additionally, Sphinx documentation can be generated from the OpenAPI specification with `tox -e openapi-doc` .

The Ceph RESTful OpenAPI specification is dynamically generated from the `Controllers` in `controllers/` directory. However, by default it is not very detailed, so there are two decorators that can and should be used to add more information:

- `@EndpointDoc()` for documentation of endpoints. It has four optional arguments (explained below): `description` , `group` , `parameters` and `responses` .
- `@ControllerDoc()` for documentation of controller or group associated with the endpoints. It only takes the two first arguments: `description` and `group` .

`description` : A a string with a short (1-2 sentences) description of the object.

`group` : By default, an endpoint is grouped together with other endpoints within the same controller class. `group` is a string that can be used to assign an endpoint or all endpoints in a class to another controller or a conceived group name.

`parameters` : A dict used to describe path, query or request body parameters. By default, all parameters for an endpoint are listed on the Swagger UI page, including

information of whether the parameter is optional/required and default values. However, there will be no description of the parameter and the parameter type will only be displayed in some cases. When adding information, each parameters should be described as in the example below. Note that the parameter type should be expressed as a built-in python type and not as a string. Allowed values are `str` , `int` , `bool` , `float` .

```
1. @EndpointDoc(parameters={'my_string': (str, 'Description of my_string'))})
2. def method(my_string): pass
```

For body parameters, more complex cases are possible. If the parameter is a dictionary, the type should be replaced with a `dict` containing its nested parameters. When describing nested parameters, the same format as other parameters is used. However, all nested parameters are set as required by default. If the nested parameter is optional this must be specified as for `item2` in the example below. If a nested parameters is set to optional, it is also possible to specify the default value (this will not be provided automatically for nested parameters).

```
1. @EndpointDoc(parameters={
2.     'my_dictionary': ({
3.         'item1': (str, 'Description of item1'),
4.         'item2': (str, 'Description of item2', True), # item2 is optional
5.         'item3': (str, 'Description of item3', True, 'foo'), # item3 is optional with 'foo' as default value
6.     }, 'Description of my_dictionary'))}
7. def method(my_dictionary): pass
```

If the parameter is a `list` of primitive types, the type should be surrounded with square brackets.

```
1. @EndpointDoc(parameters={'my_list': ([int], 'Description of my_list'))})
2. def method(my_list): pass
```

If the parameter is a `list` with nested parameters, the nested parameters should be placed in a dictionary and surrounded with square brackets.

```
1. @EndpointDoc(parameters={
2.     'my_list': ([{
3.         'list_item': (str, 'Description of list_item'),
4.         'list_item2': (str, 'Description of list_item2')
5.     }], 'Description of my_list'))}
6. def method(my_list): pass
```

`responses` : A dict used for describing responses. Rules for describing responses are the same as for request body parameters, with one difference: responses also needs to be assigned to the related response code as in the example below:

```
1. @EndpointDoc(responses={
2.     '400': {'my_response': (str, 'Description of my_response'))})
```

```
3. def method(): pass
```

## Error Handling in Python

Good error handling is a key requirement in creating a good user experience and providing a good API.

Dashboard code should not duplicate C++ code. Thus, if error handling in C++ is sufficient to provide good feedback, a new wrapper to catch these errors is not necessary. On the other hand, input validation is the best place to catch errors and generate the best error messages. If required, generate errors as soon as possible.

The backend provides few standard ways of returning errors.

First, there is a generic Internal Server Error:

```
1. Status Code: 500
2. {
3.     "version": <cherrypy version, e.g. 13.1.0>,
4.     "detail": "The server encountered an unexpected condition which prevented it from fulfilling the
5.     request.",
5. }
```

For errors generated by the backend, we provide a standard error format:

```
1. Status Code: 400
2. {
3.     "detail": str(e),      # E.g. "[errno -42] <some error message>"
4.     "component": "rbd",   # this can be null to represent a global error code
5.     "code": "3",          # Or a error name, e.g. "code": "some_error_key"
6. }
```

In case, the API Endpoints uses @ViewCache to temporarily cache results, the error looks like so:

```
1. Status Code 400
2. {
3.     "detail": str(e),      # E.g. "[errno -42] <some error message>"
4.     "component": "rbd",   # this can be null to represent a global error code
5.     "code": "3",          # Or a error name, e.g. "code": "some_error_key"
6.     'status': 3,          # Indicating the @ViewCache error status
7. }
```

In case, the API Endpoints uses a task the error looks like so:

```
1. Status Code 400
2. {
3.     "detail": str(e),      # E.g. "[errno -42] <some error message>"
4.     "component": "rbd",   # this can be null to represent a global error code
```

```

5.     "code": "3",           # Or a error name, e.g. "code": "some_error_key"
6.     "task": {              # Information about the task itself
7.       "name": "taskname",
8.       "metadata": {...}
9.     }
10. }

```

Our WebUI should show errors generated by the API to the user. Especially field-related errors in wizards and dialogs or show non-intrusive notifications.

Handling exceptions in Python should be an exception. In general, we should have few exception handlers in our project. Per default, propagate errors to the API, as it will take care of all exceptions anyway. In general, log the exception by adding `logger.exception()` with a description to the handler.

We need to distinguish between user errors from internal errors and programming errors. Using different exception types will ease the task for the API layer and for the user interface:

Standard Python errors, like `SystemError`, `ValueError` or `KeyError` will end up as internal server errors in the API.

In general, do not `return` error responses in the REST API. They will be returned by the error handler. Instead, raise the appropriate exception.

## Plug-ins

New functionality can be provided by means of a plug-in architecture. Among the benefits this approach brings in, loosely coupled development is one of the most notable. As the Ceph Dashboard grows in feature richness, its code-base becomes more and more complex. The hook-based nature of a plug-in architecture allows to extend functionality in a controlled manner, and isolate the scope of the changes.

Ceph Dashboard relies on [Pluggy](#) to provide for plug-ing support. On top of pluggy, an interface-based approach has been implemented, with some safety checks (method override and abstract method checks).

In order to create a new plugin, the following steps are required:

1. Add a new file under `src/pybind/mgr/dashboard/plugins`.
2. Import the `PLUGIN_MANAGER` instance and the `Interfaces`.
3. Create a class extending the desired interfaces. The plug-in library will check if all the methods of the interfaces have been properly overridden.
4. Register the plugin in the `PLUGIN_MANAGER` instance.
5. Import the plug-in from within the Ceph Dashboard `module.py` (currently no dynamic loading is implemented).

The available Mixins (helpers) are:

- `CanMgr` : provides the plug-in with access to the `mgr` instance under `self.mgr` .

The available Interfaces are:

- `Initializable` : requires overriding `init()` hook. This method is run at the very beginning of the dashboard module, right after all imports have been performed.
- `Setupable` : requires overriding `setup()` hook. This method is run in the Ceph Dashboard `serve()` method, right after CherryPy has been configured, but before it is started. It's a placeholder for the plug-in initialization logic.
- `HasOptions` : requires overriding `get_options()` hook by returning a list of `Options()` . The options returned here are added to the `MODULE_OPTIONS` .
- `HasCommands` : requires overriding `register_commands()` hook by defining the commands the plug-in can handle and decorating them with `@CLICommand` . The commands can be optionally returned, so that they can be invoked externally (which makes unit testing easier).
- `HasControllers` : requires overriding `get_controllers()` hook by defining and returning the controllers as usual.
- `FilterRequest.BeforeHandler` : requires overriding `filter_request_before_handler()` hook. This method receives a `cherrypy.request` object for processing. A usual implementation of this method will allow some requests to pass or will raise a `cherrypy.HTTPError` based on the `request` metadata and other conditions.

New interfaces and hooks should be added as soon as they are required to implement new functionality. The above list only comprises the hooks needed for the existing plugins.

A sample plugin implementation would look like this:

```

1. # src/pybind/mgr/dashboard/plugins/mute.py
2.
3. from . import PLUGIN_MANAGER as PM
4. from . import interfaces as I
5.
6. from mgr_module import CLICommand, Option
7. import cherrypy
8.
9. @PM.add_plugin
10. class Mute(I.CanMgr, I.Setupable, I.HasOptions, I.HasCommands,
11.             I.FilterRequest.BeforeHandler, I.HasControllers):
12.     @PM.add_hook
13.     def get_options(self):
14.         return [Option('mute', default=False, type='bool')]
15.
16.     @PM.add_hook

```

```

17.     def setup(self):
18.         self.mute = self.mgr.get_module_option('mute')
19.
20.     @PM.add_hook
21.     def register_commands(self):
22.         @CLICommand("dashboard mute")
23.         def _(mgr):
24.             self.mute = True
25.             self.mgr.set_module_option('mute', True)
26.             return 0
27.
28.     @PM.add_hook
29.     def filter_request_before_handler(self, request):
30.         if self.mute:
31.             raise cherrypy.HTTPError(500, "I'm muted :-x")
32.
33.     @PM.add_hook
34.     def get_controllers(self):
35.         from ..controllers import ApiController, RESTController
36.
37.         @ApiController('/mute')
38.         class MuteController(RESTController):
39.             def get(_):
40.                 return self.mute
41.
42.         return [MuteController]

```

Additionally, a helper for creating plugins `SimplePlugin` is provided. It facilitates the basic tasks (Options, Commands, and common Mixins). The previous plugin could be rewritten like this:

```

1. from . import PLUGIN_MANAGER as PM
2. from . import interfaces as I
3. from .plugin import SimplePlugin as SP
4.
5. import cherrypy
6.
7. @PM.add_plugin
8. class Mute(SP, I.Setupable, I.FilterRequest.BeforeHandler, I.HasControllers):
9.     OPTIONS = [
10.         SP.Option('mute', default=False, type='bool')
11.     ]
12.
13.     def shut_up(self):
14.         self.set_option('mute', True)
15.         self.mute = True
16.         return 0
17.
18.     COMMANDS = [
19.         SP.Command("dashboard mute", handler=shut_up)
20.     ]
21.

```

```
22. @PM.add_hook
23.     def setup(self):
24.         self.mute = self.get_option('mute')
25.
26.     @PM.add_hook
27.     def filter_request_before_handler(self, request):
28.         if self.mute:
29.             raise cherrypy.HTTPError(500, "I'm muted :-x")
30.
31.     @PM.add_hook
32.     def get_controllers(self):
33.         from ..controllers import ApiController, RESTController
34.
35.         @ApiController('/mute')
36.         class MuteController(RESTController):
37.             def get(_):
38.                 return self.mute
39.
40.         return [MuteController]
```

# Ceph Dashboard Design Goals

---

## Note

this document is intended to provide a focal point for discussing the overall design principles for mgr/dashboard

## Introduction

---

Most distributed storage architectures are inherently complex, and can present a management challenge to Operations teams who are typically stretched across multiple product and platform disciplines. In general terms, the complexity of any solution can have a direct bearing on the operational costs incurred to manage it. The answer is simple...make it simple :)

This document is intended to highlight Ceph Dashboard design goals which may help to

- reduce complexity
- increase productivity
- improve time-to-value
- increase observability

## Understanding the Persona of the Target User

---

Ceph has historically been administered from the CLI. The CLI has always and will always offer the richest, most flexible way to install and manage a Ceph cluster. Administrators who require and demand this level of control are unlikely to adopt a UI for anything more than a technical curiosuty.

The relevance of the UI is therefore more critical for a new SysAdmin, where it can help technology adoption and reduce the operational friction that is normally experienced when implementing a new solution.

Understanding the target user persona is therefore a fundamental first step in design. Attempting to design a UI that meets the requirements of a 'seasoned' Ceph Administrator or Developer, and a relatively new SysAdmin is unlikely to satisfy either user group.

## Design Principles

---

## Key Principles

1. **Clarity and consistency.** The UI should ensure the data shown is unambiguous and consistent across different views
2. **Data timeliness.** Data displayed in the UI must be timely. State information **must** be reasonably recent for it to be relevant and acted upon with confidence. In addition, the age of the data should be shown as an age (e.g. 20s ago) rather than UTC timestamps to make it more immediately consumable by the Administrator.
3. **Automate through workflows.** If the admin has to follow a ‘recipe’ to perform a task, the goal of the dashboard UI should be to implement the flow.
4. **Provide a natural next step.** The UI **is** the expert system, so instead of expecting the user to know where to they go next, the UI should lead them. This means linking components together to establish a flow, and deeper integration between the alertmanager implementation and the dashboard elements enabling an Admin to efficiently step from alert to affected component.
5. **Platform visibility.** The platform (OS and hardware configuration) is a fundamental component of the solution, so providing platform level insights can help deliver a more holistic view of the Ceph cluster.
6. **Jargon Busting.** Jargon is an unavoidable component of most systems. However, a good system will include inline help to support new and infrequent users of the UI.

## Common Pitfalls

- Don’t re-implement CLI commands in the UI. The sysadmin will likely use the CLI primitives in scripts to automate tasks, so by simply adding a CLI feature we miss the workflow and add complexity, which potentially ‘bloats’ the UI.
- Don’t think like a developer...try and adopt the mindset of an Administrator, who only works with the Ceph cluster part-time - this is the reality for today’s Operations teams.

## Focus On User Experience

---

Ultimately, the goal must be to move away from pushing complexity onto the GUI user through multi-step workflows like iSCSI configuration, or setting specific cluster flags in defined sequences. Simplicity, should be the goal for the UI...let’s leave complexity to the CLI.

# Ceph Internals

## Note

If you're looking for how to use Ceph as a library from your own software, please see [API Documentation](#).

You can start a development mode Ceph cluster, after compiling the source, with:

```
1. cd build  
2. OSD=3 MON=3 MGR=3 ./src/vstart.sh -n -x  
3. # check that it's there  
4. bin/ceph health
```

## Mailing list

The `dev@ceph.io` list is for discussion about the development of Ceph, its interoperability with other technology, and the operations of the project itself. Subscribe by sending a message to `dev-request@ceph.io` with the line:

```
1. subscribe ceph-devel
```

in the body of the message.

The `ceph-devel@vger.kernel.org` list is for discussion and patch review for the Linux kernel Ceph client component. Subscribe by sending a message to `majordomo@vger.kernel.org` with the line:

```
1. subscribe ceph-devel
```

in the body of the message.

# Governance

The Ceph open source community is guided by a few different groups.

## Project Leader

The Ceph project is currently led by Sage Weil <[sage@redhat.com](mailto:sage@redhat.com)>. The project leader is responsible for guiding the overall direction of the project and ensuring that the developer and user communities are healthy.

## Committers

Committers are project contributors who have write access to the central Ceph code repositories, currently hosted on GitHub. This group of developers is collectively empowered to make changes to the Ceph source code.

Generally speaking, no individual should make a change in isolation: all code contributions go through a collaborative review process (and undergo testing) before being merged. The specifics of this process are dynamic and evolving over time.

New committers are added to the project (or committers removed from the project) at the discretion of the Ceph Leadership Team (below). The criteria for becoming a contributor include a consistent level of quality and engagement in the project over time.

## Ceph Leadership Team

The Ceph Leadership Team (CLT) is a collection of component leads and other core developers who collectively make technical decisions for the project. These decisions are generally made by consensus, although voting may be used if necessary.

The CLT meets weekly via video chat to discuss any pending issues or decisions. Minutes for the CLT meetings are published at <https://pad.ceph.com/p/clt-weekly-minutes>.

Committers are added to or removed from the CLT at the discretion of the CLT itself.

Current CLT members are:

- Abhishek Lekshmanan <[abhishek@suse.com](mailto:abhishek@suse.com)>
- Casey Bodley <[cbodley@redhat.com](mailto:cbodley@redhat.com)>
- Ernesto Puerta <[epuerta@redhat.com](mailto:epuerta@redhat.com)>

- Gregory Farnum <[gfarnum@redhat.com](mailto:gfarnum@redhat.com)>
- Haomai Wang <[haomai@xsky.com](mailto:haomai@xsky.com)>
- Jason Dillaman <[dillaman@redhat.com](mailto:dillaman@redhat.com)>
- Josh Durgin <[jdurgin@redhat.com](mailto:jdurgin@redhat.com)>
- João Eduardo Luis <[joao@suse.de](mailto:joao@suse.de)>
- Ken Dreyer <[kdreyer@redhat.com](mailto:kdreyer@redhat.com)>
- Matt Benjamin <[mbenjami@redhat.com](mailto:mbenjami@redhat.com)>
- Myoungwon Oh <[omwmw@sk.com](mailto:omwmw@sk.com)>
- Neha Ojha <[nojha@redhat.com](mailto:nojha@redhat.com)>
- Patrick Donnelly <[pdonnell@redhat.com](mailto:pdonnell@redhat.com)>
- Sage Weil <[sage@redhat.com](mailto:sage@redhat.com)>
- Sebastian Wagner <[swagner@suse.com](mailto:swagner@suse.com)>
- Xie Xingguo <[xie.xingguo@zte.com.cn](mailto:xie.xingguo@zte.com.cn)>
- Yehuda Sadeh <[yehuda@redhat.com](mailto:yehuda@redhat.com)>

## Component Leads

Each major subcomponent of the Ceph project has a lead engineer who is responsible for guiding and coordinating development. The leads are nominated or appointed at the discretion of the project leader or the CLT. Leads responsibilities include:

- guiding the (usually) daily “stand-up” coordination calls over video chat
- building the development roadmap for each release cycle
- coordinating development activity between contributors
- ensuring that contributions are reviewed
- ensuring that different proposed changes do not conflict
- ensuring that testing remains robust (new features include tests, changes do not break tests, etc.)

All component leads are included on the CLT. They are expected to report progress and status updates to the rest of the leadership team and to help facilitate any cross-component coordination of development.

## The Ceph Foundation

The Ceph Foundation is organized as a directed fund under the Linux Foundation and is tasked with supporting the Ceph project community and ecosystem. It has no direct

control over the technical direction of the Ceph open source project beyond offering feedback and input into the collaborative development process.

For more information, see [Ceph Foundation](#).

# Ceph Foundation

---

The Ceph Foundation exists to enable industry members to collaborate and pool resources to support the Ceph project community. The Foundation provides an open, collaborative, and neutral home for project stakeholders to coordinate their development and community investments in the Ceph ecosystem.

The Ceph Foundation is organized as a directed fund under the Linux Foundation. Premier and General Member organizations contribute a yearly fee to become members. Associate members are educational institutions or government organizations and are invited to join at no cost.

For more information, see <https://ceph.com/foundation>.

## Members

---

### Premier

- [Amihan](#)
- [Bloomberg](#)
- [Canonical](#)
- [China Mobile](#)
- [DigitalOcean](#)
- [Intel](#)
- [OVH](#)
- [Red Hat](#)
- [Samsung Electronics](#)
- [SoftIron](#)
- [SUSE](#)
- [Western Digital](#)
- [XSKY](#)
- [ZTE](#)

### General

- [ARM](#)
- [Catalyst Cloud](#)
- [Cloudbase Solutions](#)
- [Clyso](#)
- [croit](#)
- [DiDi](#)
- [EasyStack](#)
- [ISS](#)
- [QCT](#)
- [SinoRail](#)
- [Vexxhost](#)

## Associate

- [Boston University](#)
- [Center for Research in Open Source Systems \(CROSS\)](#)
- [CERN](#)
- [FASRC](#)
- [grnet](#)
- [Monash University](#)
- [NRF SARAO](#)
- [Science & Technology Facilities Council \(STFC\)](#)
- [University of Michigan](#)
- [SWITCH](#)

## Governing Board

---

The Governing Board consists of all Premier members, a representative for the General members, a representative for the Associate members, and a representative from the Ceph Leadership Team (the technical governance body). The board is responsible for:

- Building and approving an annual budget for spending in support of the Ceph project

- Establishing ad-hoc committees to address current needs of the project
- Coordinating outreach or marketing
- Meeting regularly to discuss Foundation activities, the status of the Ceph project, and overall project strategy
- Voting on any decisions or matters before the board

The Ceph Foundation board is not responsible for and does not have any direct control over the technical governance of Ceph. Development and engineering activities are managed through traditional open source processes and are overseen by the [Ceph Leadership Team](#). For more information see [Governance](#).

## Members

- Anjaneya “Reddy” Chagam (Intel)
- Dan van der Ster (CERN) - Associate member representative
- Haomai Wang (XSKY)
- James Page (Canonical)
- Lenz Grimmer (SUSE) - Ceph Leadership Team representative
- Lars Marowsky-Bree (SUSE)
- Matias Bjorling (Western Digital)
- Matthew Leonard (Bloomberg)
- Mike Perez (Red Hat) - Ceph community manager
- Myoungwon Oh (Samsung Electronics)
- Paul Emmerich (croit) - General member representative
- Paweł Sadowski (OVH)
- Phil Straw (SoftIron)
- Robin Johnson (DigitalOcean)
- Sage Weil (Red Hat) - Ceph project leader
- Winston Damarillio (Amihan)
- Xie Xingguo (ZTE)
- Zhang Shaowen (China Mobile)

# Joining

---

For information about joining the Ceph Foundation, please contact [membership@linuxfoundation.org](mailto:membership@linuxfoundation.org).

# ceph-volume

Deploy OSDs with different device technologies like lvm or physical disks using pluggable tools (`lvm` itself is treated like a plugin) and trying to follow a predictable, and robust way of preparing, activating, and starting OSDs.

[Overview](#) | [Plugin Guide](#) |

## Command Line Subcommands

There is currently support for `lvm`, and plain disks (with GPT partitions) that may have been deployed with `ceph-disk`.

`zfs` support is available for running a FreeBSD cluster.

- `lvm`
- `simple`
- `zfs`

## Node inventory

The `inventory` subcommand provides information and metadata about a nodes physical disk inventory.

# Migrating

Starting on Ceph version 13.0.0, `ceph-disk` is deprecated. Deprecation warnings will show up that will link to this page. It is strongly suggested that users start consuming `ceph-volume`. There are two paths for migrating:

1. Keep OSDs deployed with `ceph-disk`: The `simple` command provides a way to take over the management while disabling `ceph-disk` triggers.
2. Redeploy existing OSDs with `ceph-volume`: This is covered in depth on [Replacing an OSD](#)

For details on why `ceph-disk` was removed please see the [Why was ceph-disk replaced?](#) section.

# New deployments

For new deployments, `lvm` is recommended, it can use any logical volume as input for data OSDs, or it can setup a minimal/naive logical volume from a device.

# Existing OSDs

If the cluster has OSDs that were provisioned with `ceph-disk`, then `ceph-volume` can take over the management of these with `simple`. A scan is done on the data device or OSD directory, and `ceph-disk` is fully disabled. Encryption is fully supported.

# Ceph Releases (general)

## Active stable releases

Version	Initial release	Latest	End of life (estimated)
octopus	Mar 2020	15.2.7	2022-06-01
nautilus	Mar 2019	14.2.14	2021-06-01

## Understanding the release cycle

Starting with the Nautilus release (14.2.0), there is a new stable release cycle every year, targeting March 1st. Each stable release series will receive a name (e.g., 'Mimic') and a major release number (e.g., 13 for Mimic because 'M' is the 13th letter of the alphabet).

Releases are named after a species of cephalopod (usually the common name, since the latin names are harder to remember or pronounce).

Version numbers have three components,  $x.y.z$ .  $x$  identifies the release cycle (e.g., 13 for Mimic).  $y$  identifies the release type:

- $x.0.z$  - development releases (for early testers and the brave at heart)
- $x.1.z$  - release candidates (for test clusters, brave users)
- $x.2.z$  - stable/bugfix releases (for users)

This versioning convention started with the 9.y.z Infernalis cycle. Prior to that, versions looked with 0.y for development releases and 0.y.z for stable series.

## Development releases ( $x.0.z$ )

Each development release ( $x.0.z$ ) freezes the master development branch and applies [integration and upgrade tests](#) before it is released. Once released, there is no effort to backport fixes; developer focus is on the next development release which is usually only a few weeks away.

- Development release every 8 to 12 weeks
- Intended for testing, not production deployments
- Full integration testing

- Upgrade testing from the last stable release(s)
- Every effort is made to allow *offline* upgrades from previous development releases (meaning you can stop all daemons, upgrade, and restart). No attempt is made to support online rolling upgrades between development releases. This facilitates deployment of development releases on non-production test clusters without repopulating them with data.

## Release candidates (x.1.z)

There is a feature release roughly eight (8) weeks prior to the planned initial stable release, after which focus shifts to stabilization and bug fixes only.

- Release candidate release every 1-2 weeks
- Intended for final testing and validation of the upcoming stable release

## Stable releases (x.2.z)

Once the initial stable release is made (x.2.0), there are semi-regular bug-fix point releases with bug fixes and (occasionally) small feature backports. Bug fixes are accumulated and included in the next point release.

- Stable point release every 4 to 6 weeks
- Intended for production deployments
- Bug fix backports for two full release cycles.
- Online, rolling upgrade support and testing from the last two (2) stable release(s) (starting from Luminous).
- Online, rolling upgrade support and testing from prior stable point releases

For each stable release:

- **Integration and upgrade tests** are run on a regular basis and **their results** analyzed by Ceph developers.
- **Issues** fixed in the development branch (master) are scheduled to be backported.
- When an issue found in the stable release is **reported**, it is triaged by Ceph developers.
- The **stable releases and backport team** publishes **point releases** including fixes that have been backported to the stable release.

## Lifetime of stable releases

---

The lifetime of a stable release series is calculated to be approximately 24 months (i.e., two 12 month release cycles) after the month of the first release. For example, Mimic (13.2.z) will reach end of life (EOL) shortly after Octopus (15.2.0) is released. The lifetime of a release may vary because it depends on how quickly the stable releases are published.

In the case of Jewel and Kraken, the lifetime was slightly different than described above. Prior to Luminous, only every other stable release was an “LTS” release. Therefore,

- Upgrade scenarios “Jewel -> Kraken -> Luminous” and “Jewel -> Luminous” were expected to work.
- Upgrades from Jewel or Kraken must upgrade to Luminous first before proceeding further (e.g., Kraken -> Luminous -> Mimic but not Kraken -> Mimic).
- Jewel was maintained until Mimic was released (June 2018).
- Kraken is no longer maintained.

Detailed information on all releases, past and present, can be found at [Ceph Releases \(index\)](#)

## Release timeline

Date	development	octopus	nautilus	mimic	luminous	kraken	jewel	i
Nov 2020	-	15.2.7	-	-	-	-	-	-
Nov 2020	-	15.2.6	-	-	-	-	-	-
Nov 2020	-	-	14.2.14	-	-	-	-	-
Nov 2020	-	-	14.2.13	-	-	-	-	-
Sep 2020	-	-	14.2.12	-	-	-	-	-
Sep 2020	-	15.2.5	-	-	-	-	-	-
Aug 2020	-	-	14.2.11	-	-	-	-	-
Jun 2020	-	15.2.4	-	-	-	-	-	-

Jun 2020	-	-	14.2.10	-	-	-	-	-
May 2020	-	15.2.3	-	-	-	-	-	-
May 2020	-	15.2.2	-	-	-	-	-	-
Apr 2020	-	-	-	13.2.10	-	-	-	-
Apr 2020	-	-	-	13.2.9	-	-	-	-
Apr 2020	-	-	14.2.9	-	-	-	-	-
Apr 2020	-	15.2.1	-	-	-	-	-	-
Mar 2020	-	15.2.0	-	-	-	-	-	-
Mar 2020	15.1.1	-	-	-	-	-	-	-
Mar 2020	-	-	14.2.8	-	-	-	-	-
Jan 2020	-	-	14.2.7	-	-	-	-	-
Jan 2020	-	-	-	-	12.2.13	-	-	-
Jan 2020	15.1.0	-	-	-	-	-	-	-
Jan 2020	-	-	14.2.6	-	-	-	-	-
Dec 2019	-	-	-	13.2.8	-	-	-	-
Dec 2019	-	-	14.2.5	-	-	-	-	-
Nov 2019	-	-	-	13.2.7	-	-	-	-
Sep								

2019	-	-	<a href="#">14.2.4</a>	-	-	-	-	-
Sep 2019	-	-	<a href="#">14.2.3</a>	-	-	-	-	-
Jul 2019	-	-	<a href="#">14.2.2</a>	-	-	-	-	-
Jun 2019	-	-	-	<a href="#">13.2.6</a>	-	-	-	-
Apr 2019	-	-	<a href="#">14.2.1</a>	-	-	-	-	-
Apr 2019	-	-	-	-	<a href="#">12.2.12</a>	-	-	-
Apr 2019	<a href="#">15.0.0</a>	-	-	-	-	-	-	-
Mar 2019	-	-	<a href="#">14.2.0</a>	-	-	-	-	-
Mar 2019	-	-	-	<a href="#">13.2.5</a>	-	-	-	-
Mar 2019	<a href="#">14.1.1</a>	-	-	-	-	-	-	-
Feb 2019	<a href="#">14.1.0</a>	-	-	-	-	-	-	-
Jan 2019	-	-	-	-	<a href="#">12.2.11</a>	-	-	-
Jan 2019	-	-	-	<a href="#">13.2.4</a>	-	-	-	-
Jan 2019	-	-	-	<a href="#">13.2.3</a>	-	-	-	-
Nov 2018	-	-	-	-	<a href="#">12.2.10</a>	-	-	-
Nov 2018	<a href="#">14.0.1</a>	-	-	-	-	-	-	-
Nov 2018	-	-	-	-	<a href="#">12.2.9</a>	-	-	-
Sep 2018	-	-	-	<a href="#">13.2.2</a>	-	-	-	-

Sep 2018	-	-	-	-	<a href="#">12.2.8</a>	-	-	-
Jul 2018	-	-	-	-	-	-	<a href="#">10.2.11</a>	-
Jul 2018	-	-	-	<a href="#">13.2.1</a>	-	-	-	-
Jul 2018	-	-	-	-	<a href="#">12.2.7</a>	-	-	-
Jul 2018	-	-	-	-	<a href="#">12.2.6</a>	-	-	-
Jun 2018	-	-	-	<a href="#">13.2.0</a>	-	-	-	-
May 2018	<a href="#">14.0.0</a>	-	-	-	-	-	-	-
May 2018	<a href="#">13.1.0</a>	-	-	-	-	-	-	-
Apr 2018	-	-	-	-	<a href="#">12.2.5</a>	-	-	-
Apr 2018	<a href="#">13.0.2</a>	-	-	-	-	-	-	-
Feb 2018	-	-	-	-	<a href="#">12.2.4</a>	-	-	-
Feb 2018	-	-	-	-	<a href="#">12.2.3</a>	-	-	-
Feb 2018	<a href="#">13.0.1</a>	-	-	-	-	-	-	-
Dec 2017	-	-	-	-	<a href="#">12.2.2</a>	-	-	-
Oct 2017	-	-	-	-	-	-	<a href="#">10.2.10</a>	-
Sep 2017	-	-	-	-	<a href="#">12.2.1</a>	-	-	-
Aug 2017	-	-	-	-	<a href="#">12.2.0</a>	-	-	-
Aug	-	-	-	-	-	<a href="#">11.2.1</a>	-	-

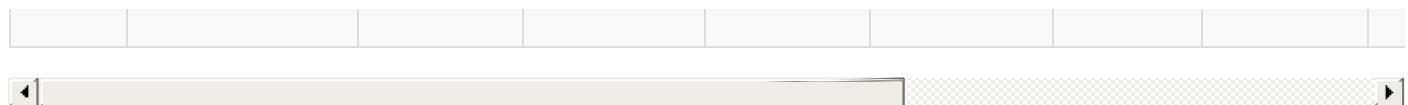
Aug 2017	-	-	-	-	-	-	<a href="#">11.2.1</a>	-	-
Aug 2017	13.0.0	-	-	-	-	-	-	-	-
Aug 2017	12.1.4	-	-	-	-	-	-	-	-
Aug 2017	12.1.3	-	-	-	-	-	-	-	-
Aug 2017	12.1.2	-	-	-	-	-	-	-	-
Jul 2017	-	-	-	-	-	-	-	<a href="#">10.2.9</a>	-
Jul 2017	-	-	-	-	-	-	-	<a href="#">10.2.8</a>	-
Jul 2017	12.1.1	-	-	-	-	-	-	-	-
Jun 2017	12.1.0	-	-	-	-	-	-	-	-
May 2017	12.0.3	-	-	-	-	-	-	-	-
Apr 2017	-	-	-	-	-	-	-	<a href="#">10.2.7</a>	-
Apr 2017	12.0.2	-	-	-	-	-	-	-	-
Mar 2017	-	-	-	-	-	-	-	<a href="#">10.2.6</a>	-
Mar 2017	12.0.1	-	-	-	-	-	-	-	-
Feb 2017	-	-	-	-	-	-	-	-	-
Feb 2017	12.0.0	-	-	-	-	-	-	-	-
Jan 2017	-	-	-	-	-	-	<a href="#">11.2.0</a>	-	-
Jan 2017	11.1.1	-	-	-	-	-	-	-	-

Dec 2016	-	-	-	-	-	-	-	<a href="#">10.2.5</a>	-
Dec 2016	-	-	-	-	-	-	-	<a href="#">10.2.4</a>	-
Dec 2016	<a href="#">11.1.0</a>	-	-	-	-	-	-	-	-
Oct 2016	<a href="#">11.0.2</a>	-	-	-	-	-	-	-	-
Oct 2016	<a href="#">11.0.1</a>	-	-	-	-	-	-	-	-
Sep 2016	-	-	-	-	-	-	-	<a href="#">10.2.3</a>	-
Aug 2016	-	-	-	-	-	-	-	-	-
Aug 2016	-	-	-	-	-	-	-	-	-
Jun 2016	-	-	-	-	-	-	-	<a href="#">10.2.2</a>	-
Jun 2016	<a href="#">11.0.0</a>	-	-	-	-	-	-	-	-
May 2016	-	-	-	-	-	-	-	<a href="#">10.2.1</a>	-
May 2016	-	-	-	-	-	-	-	-	-
Apr 2016	-	-	-	-	-	-	-	<a href="#">10.2.0</a>	-
Apr 2016	<a href="#">10.1.2</a>	-	-	-	-	-	-	-	-
Apr 2016	<a href="#">10.1.1</a>	-	-	-	-	-	-	-	-
Mar 2016	<a href="#">10.1.0</a>	-	-	-	-	-	-	-	-
Mar 2016	<a href="#">10.0.5</a>	-	-	-	-	-	-	-	-

2016	<a href="#">10.0.4</a>	-	-	-	-	-	-	-	-	-
Feb 2016	-	-	-	-	-	-	-	-	-	<a href="#">C</a>
Feb 2016	-	-	-	-	-	-	-	-	-	-
Feb 2016	<a href="#">10.0.3</a>	-	-	-	-	-	-	-	-	-
Jan 2016	<a href="#">10.0.2</a>	-	-	-	-	-	-	-	-	-
Dec 2015	<a href="#">10.0.1</a>	-	-	-	-	-	-	-	-	-
Nov 2015	-	-	-	-	-	-	-	-	-	<a href="#">G</a>
Nov 2015	-	-	-	-	-	-	-	-	-	-
Nov 2015	<a href="#">10.0.0</a>	-	-	-	-	-	-	-	-	-
Oct 2015	-	-	-	-	-	-	-	-	-	-
Oct 2015	-	-	-	-	-	-	-	-	-	-
Oct 2015	<a href="#">9.1.0</a>	-	-	-	-	-	-	-	-	-
Aug 2015	-	-	-	-	-	-	-	-	-	-
Aug 2015	<a href="#">9.0.3</a>	-	-	-	-	-	-	-	-	-
Jul 2015	-	-	-	-	-	-	-	-	-	-
Jul 2015	<a href="#">9.0.2</a>	-	-	-	-	-	-	-	-	-
Jun 2015	-	-	-	-	-	-	-	-	-	-
Jun 2015	<a href="#">9.0.1</a>	-	-	-	-	-	-	-	-	-

May 2015	9.0.0	-	-	-	-	-	-	-	-
Apr 2015	-	-	-	-	-	-	-	-	-
Apr 2015	-	-	-	-	-	-	-	-	-
Apr 2015	-	-	-	-	-	-	-	-	-
Mar 2015	-	-	-	-	-	-	-	-	-
Feb 2015	-	-	-	-	-	-	-	-	-
Feb 2015	0.93	-	-	-	-	-	-	-	-
Feb 2015	0.92	-	-	-	-	-	-	-	-
Jan 2015	-	-	-	-	-	-	-	-	-
Jan 2015	0.91	-	-	-	-	-	-	-	-
Dec 2014	0.90	-	-	-	-	-	-	-	-
Dec 2014	0.89	-	-	-	-	-	-	-	-
Nov 2014	0.88	-	-	-	-	-	-	-	-
Oct 2014	-	-	-	-	-	-	-	-	-
Oct 2014	-	-	-	-	-	-	-	-	-
Oct 2014	-	-	-	-	-	-	-	-	-
Oct 2014	0.86	-	-	-	-	-	-	-	-

Sep 2014	0.85	-	-	-	-	-	-	-	-
Aug 2014	0.84	-	-	-	-	-	-	-	-
Jul 2014	-	-	-	-	-	-	-	-	-
Jul 2014	-	-	-	-	-	-	-	-	-
Jul 2014	-	-	-	-	-	-	-	-	-
Jul 2014	-	-	-	-	-	-	-	-	-
Jul 2014	0.83	-	-	-	-	-	-	-	-
Jun 2014	0.82	-	-	-	-	-	-	-	-
Jun 2014	0.81	-	-	-	-	-	-	-	-
May 2014	-	-	-	-	-	-	-	-	-
May 2014	-	-	-	-	-	-	-	-	-
Apr 2014	0.79	-	-	-	-	-	-	-	-
Mar 2014	0.78	-	-	-	-	-	-	-	-
Feb 2014	0.77	-	-	-	-	-	-	-	-
Jan 2014	0.76	-	-	-	-	-	-	-	-
Jan 2014	0.75	-	-	-	-	-	-	-	-
Dec 2013	0.74	-	-	-	-	-	-	-	-
Dec 2013	0.73	-	-	-	-	-	-	-	-



# Ceph Releases (index)

---

## Active Releases

---

- [v15.2.7 Octopus](#)
- [v15.2.6 Octopus](#)
- [v15.2.5 Octopus](#)
- [v15.2.4 Octopus](#)
- [v15.2.3 Octopus](#)
- [v15.2.2 Octopus](#)
- [v15.2.1 Octopus](#)
- [v15.2.0 Octopus](#)
- [v14.2.15 Nautilus](#)
- [v14.2.14 Nautilus](#)
- [v14.2.13 Nautilus](#)
- [v14.2.12 Nautilus](#)
- [v14.2.11 Nautilus](#)
- [v14.2.10 Nautilus](#)
- [v14.2.9 Nautilus](#)
- [v14.2.8 Nautilus](#)
- [v14.2.7 Nautilus](#)
- [v14.2.6 Nautilus](#)
- [v14.2.5 Nautilus](#)
- [v14.2.4 Nautilus](#)
- [v14.2.3 Nautilus](#)
- [v14.2.2 Nautilus](#)
- [v14.2.1 Nautilus](#)
- [v14.2.0 Nautilus](#)

## Archived Releases

---

- [Archived releases index](#)

## v15.2.7 Octopus

---

This is the 7th backport release in the Octopus series. This release fixes a serious bug in RGW that has been shown to cause data loss when a read of a large RGW object (i.e., one with at least one tail segment) takes longer than one half the time specified in the configuration option `rgw_gc_obj_min_wait`. The bug causes the tail segments of that read object to be added to the RGW garbage collection queue, which will in turn cause them to be deleted after a period of time.

## Changelog

---

- rgw: during GC defer, prevent new GC enqueue ([issue#47866](#), [pr#38249](#), Eric Ivancich, Casey Bodley)

## v15.2.6 Octopus

---

This is the 6th backport release in the Octopus series. This release fixes a security flaw affecting Messenger v1 & v2. We recommend users to update to this release.

## Notable Changes

---

- CVE 2020-25660: CEPHX\_V2 replay attack protection lost, for Messenger v1 & v2 (Ilya Dryomov)

## Changelog

---

- mon/MonClient: bring back CEPHX\_V2 authorizer challenges (Ilya Dryomov)

## v15.2.5 Octopus

---

This is the fifth release of the Ceph Octopus stable release series. This release brings a range of fixes across all components. We recommend that all Octopus users upgrade to this release.

## Notable Changes

---

- CephFS: Automatic static subtree partitioning policies may now be configured using the new distributed and random ephemeral pinning extended attributes on directories. See the documentation for more information:  
<https://docs.ceph.com/docs/master/cephfs/multimds/>

- Monitors now have a config option `mon_osd_warn_num_repaired`, 10 by default. If any OSD has repaired more than this many I/O errors in stored data a `OSD_TOO_MANY_REPAIRS` health warning is generated.
- Now when noscrub and/or no deep-scrub flags are set globally or per pool, scheduled scrubs of the type disabled will be aborted. All user initiated scrubs are NOT interrupted.
- Fix an issue with osdmaps not being trimmed in a healthy cluster ([issue#47297](#), [pr#36981](#))

## Changelog

---

- bluestore,core: bluestore: blk:BlockDevice.cc: use pending\_aios instead of iovec size as ios num ([pr#36668](#), weixinwei)
- bluestore,tests: test/store\_test: refactor bluestore spillover test ([pr#34943](#), Igor Fedotov)
- bluestore,tests: tests: objectstore/store\_test: kill ExcessiveFragmentation test case ([pr#36049](#), Igor Fedotov)
- bluestore: bluestore: Rescue procedure for extremely large bluefs log ([pr#36123](#), Adam Kupczyk)
- bluestore: octopus:os/bluestore: improve/fix bluefs stats reporting ([pr#35748](#), Igor Fedotov)
- bluestore: os/bluestore: fix bluefs log growth ([pr#36621](#), Adam Kupczyk, Jianpeng Ma)
- bluestore: os/bluestore: simplify Onode pin/unpin logic ([pr#36795](#), Igor Fedotov)
- build/ops: Revert “mgr/osd\_support: remove module and all traces” ([pr#36973](#), Sebastian Wagner)
- build/ops: ceph-iscsi: selinux fixes ([pr#36302](#), Mike Christie)
- build/ops: mgr/dashboard/api: reduce amount of daemon logs ([pr#36693](#), Ernesto Puerta)
- ceph-volume: add dmcrypt support in raw mode ([pr#35830](#), Guillaume Abrioux)
- ceph-volume: add drive-group subcommand ([pr#36558](#), Jan Fajerski, Sebastian Wagner)
- ceph-volume: add tests for new functions that run LVM commands ([pr#36614](#), Rishabh Dave)
- ceph-volume: don't use container classes in api/lvm.py ([pr#35879](#), Rishabh Dave,

Guillaume Abrioux)

- ceph-volume: fix lvm functional tests ([pr#36409](#), Jan Fajerski)
- ceph-volume: handle idempotency with batch and explicit scenarios ([pr#35880](#), Andrew Schoen)
- ceph-volume: remove container classes from api/lvm.py ([pr#36608](#), Rishabh Dave)
- ceph-volume: report correct rejected reason in inventory if device type is invalid ([pr#36410](#), Satoru Takeuchi)
- ceph-volume: run flake8 in python3 ([pr#36588](#), Jan Fajerski)
- cephfs,common: common: ignore SIGHUP prior to fork ([issue#46269](#), [pr#36195](#), Willem Jan Withagen, hzwuhongsong)
- cephfs,core,mgr: mgr/status: metadata is fetched async ([pr#36630](#), Michael Fritch)
- cephfs,core,rbd,rgw: librados: add LIBRADOS\_SUPPORTS\_GETADDRS support ([pr#36643](#), Xiubo Li)
- cephfs,mgr: mgr/volumes/nfs: Add interface for adding user defined configuration ([pr#36635](#), Varsha Rao)
- cephfs,mon: mon/MDSMonitor: copy MDS info which may be removed ([pr#36035](#), Patrick Donnelly)
- cephfs,pybind: pybind/ceph\_volume\_client: Fix PEP-8 SyntaxWarning ([pr#36100](#), Đặng Minh Dũng)
- cephfs,tests: mgr/fs/volumes: misc fixes ([pr#36327](#), Patrick Donnelly, Kotresh HR)
- cephfs,tests: tests: Revert “Revert “qa/suites/rados/mgr/tasks/module\_selftest: whitelist ... ([issue#43943](#), [pr#36042](#), Venky Shankar)
- cephfs,tests: tests: qa/tasks/cephfs/cephfs\_test\_case.py: skip cleaning the core dumps when in program case ([pr#36043](#), Xiubo Li)
- cephfs,tests: tests: qa/tasks: make sh() in vstart\_runner.py identical with teuthology.orchestra.remote.sh ([pr#36044](#), Jos Collin)
- cephfs: Update nfs-ganesha package requirements doc backport ([pr#36063](#), Varsha Rao)
- cephfs: cephfs: client: fix setxattr for 0 size value (NULL value) ([pr#36045](#), Sidharth Anupkrishnan)
- cephfs: cephfs: client: fix snap directory atime ([pr#36039](#), Luis Henriques)
- cephfs: cephfs: client: release the client\_lock before copying data in read ([pr#36046](#), Chencan)

- cephfs: client: expose ceph.quota.max\_bytes xattr within snapshots ([pr#36403](#), Shyamsundar Ranganathan)
- cephfs: client: introduce timeout for client shutdown ([issue#44276](#), [pr#35962](#), "Yan, Zheng", Venky Shankar)
- cephfs: mds/MDSRank: fix typo in "unrecognized" ([pr#36197](#), Nathan Cutler)
- cephfs: mds: add ephemeral random and distributed export pins ([pr#35759](#), Patrick Donnelly, Sidharth Anupkrishnan)
- cephfs: mds: fix filelock state when Fc is issued ([pr#35842](#), Xiubo Li)
- cephfs: mds: reset heartbeat in EMetaBlob replay ([pr#36040](#), Yanhu Cao)
- cephfs: mgr/nfs: Check if pseudo path is absolute path ([pr#36299](#), Varsha Rao)
- cephfs: mgr/nfs: Update MDCACHE block in ganesha config and doc about nfs-cephadm in vstart ([pr#36224](#), Varsha Rao)
- cephfs: mgr/volumes: Deprecate protect/unprotect CLI calls for subvolume snapshots ([pr#36126](#), Shyamsundar Ranganathan)
- cephfs: mgr/volumes: fix "ceph nfs export" help messages ([pr#36220](#), Nathan Cutler)
- cephfs: nfs backport ([pr#35499](#), Jeff Layton, Varsha Rao, Ramana Raja, Kefu Chai)
- common,core: common, osd: add sanity checks around osd\_scrub\_max\_preemptions ([pr#36034](#), xie xingguo)
- common,rbd,tools: rbd: immutable-object-cache: fixed crashes on start up ([pr#36660](#), Jason Dillaman)
- common,rbd: crush/CrushWrapper: rebuild reverse maps after rebuilding crush map ([pr#36662](#), Jason Dillaman)
- common: common: log: fix timestamp precision of log can't set to millisecond ([pr#36048](#), Guan yunfei)
- core,mgr: mgr: decrease pool stats if pg was removed ([pr#36667](#), Aleksei Gutikov)
- core,rbd: osd/OSDCap: rbd profile permits use of "rbd\_info" ([pr#36414](#), Florian Florensa)
- core,tools: tools/rados: Set locator key when exporting or importing a pool ([pr#36666](#), Iain Buclaw)
- core: mon/OSDMonitor: Reset grace period if failure interval exceeds a threshold ([pr#35799](#), Sridhar Seshaayee)
- core: mon/OSDMonitor: only take in osd into consideration when trimming osd...

([pr#36981](#), Kefu Chai)

- core: mon: fix the ‘Error ERANGE’ message when conf “osd\_objectstore” is filestore ([pr#36665](#), wangyunqing)
- core: monclient: schedule first tick using mon\_client\_hunt\_interval ([pr#36633](#), Mykola Golub)
- core: osd/OSD.cc: remove osd\_lock for bench ([pr#36664](#), Neha Ojha, Adam Kupczyk)
- core: osd/PG: fix history.same\_interval\_since of merge target again ([pr#36033](#), xie xingguo)
- core: osd/PeeringState: prevent peer’s num\_objects going negative ([pr#36663](#), xie xingguo)
- core: osd/PrimaryLogPG: don’t populate watchers if replica ([pr#36029](#), Ilya Dryomov)
- core: osd: Cancel in-progress scrubs (not user requested) ([pr#36291](#), David Zafman)
- core: osd: expose osdspec\_affinity to osd\_metadata ([pr#35957](#), Joshua Schmid)
- core: osd: fix crash in \_committed\_osd\_maps if incremental osdmap crc fails ([pr#36340](#), Neha Ojha, Dan van der Ster)
- core: osd: make message cap option usable again ([pr#35737](#), Neha Ojha, Josh Durgin)
- core: osd: wakeup all threads of shard rather than one thread ([pr#36032](#), Jianpeng Ma)
- core: test: osd-backfill-stats.sh use nobackfill to avoid races in remainin... ([pr#36030](#), David Zafman)
- doc: cephadm batch backport ([pr#36450](#), Varsha Rao, Ricardo Marques, Kiefer Chang, Matthew Oliver, Paul Cuzner, Kefu Chai, Daniel-Pivonka, Sebastian Wagner, Volker Theile, Adam King, Michael Fritch, Joshua Schmid)
- doc: doc/mgr/crash: Add missing command in rm example ([pr#36690](#), Daniël Vos)
- doc: doc/rados: Fix osd\_scrub\_during\_recovery default value ([pr#36661](#), Benoît Knecht)
- doc: doc/rbd: add rbd-target-gw enable and start ([pr#36416](#), Zac Dover)
- doc: doc: PendingReleaseNotes: clean slate for 15.2.5 ([pr#35753](#), Nathan Cutler)
- mgr,pybind: pybind/mgr/balancer: use “==” and “!=” for comparing str ([pr#36036](#), Kefu Chai)

- mgr/pybind: pybind/mgr/pg\_autoscaler/module.py: do not update event if ev.pg\_num== ev.pg\_num\_target ([pr#36037](#), Neha Ojha)
- mgr/rbd: mgr/prometheus: automatically discover RBD pools for stats gathering ([pr#36411](#), Jason Dillaman)
- mgr/dashboard/api: increase API health timeout ([pr#36562](#), Ernesto Puerta)
- mgr/dashboard: Add button to copy the bootstrap token into the clipboard ([pr#35796](#), Ishan Rai)
- mgr/dashboard: Add host labels in UI ([pr#35893](#), Volker Theile)
- mgr/dashboard: Add hosts page unit tests ([pr#36350](#), Volker Theile)
- mgr/dashboard: Allow to edit iSCSI target with active session ([pr#35997](#), Ricardo Marques)
- mgr/dashboard: Always use fast angular unit tests ([pr#36267](#), Stephan Müller)
- mgr/dashboard: Configure overflow of popover in health page ([pr#36460](#), Tiago Melo)
- mgr/dashboard: Display check icon instead of true|false in various datatables ([pr#35892](#), Volker Theile)
- mgr/dashboard: Display users current bucket quota usage ([pr#35926](#), Ernesto Puerta, Avan Thakkar)
- mgr/dashboard: Extract documentation link to a component ([pr#36587](#), Tiago Melo)
- mgr/dashboard: Fix host attributes like labels are not returned ([pr#36678](#), Kiefer Chang)
- mgr/dashboard: Hide password notification when expiration date is far ([pr#35975](#), Tiago Melo)
- mgr/dashboard: Improve Summary's subscribe methods ([pr#35705](#), Tiago Melo)
- mgr/dashboard: Prometheus query error in the metrics of Pools, OSDs and RBD images ([pr#35885](#), Avan Thakkar)
- mgr/dashboard: Re-enable OSD's table autoReload ([pr#36226](#), Kiefer Chang, Tiago Melo)
- mgr/dashboard: Strange iSCSI discovery auth behavior ([pr#36782](#), Volker Theile)
- mgr/dashboard: The max. buckets field in RGW user form should be pre-filled ([pr#35795](#), Volker Theile)
- mgr/dashboard: Unable to edit iSCSI logged-in client ([pr#36611](#), Ricardo Marques)

- mgr/dashboard: Use right size in pool form ([pr#35925](#), Stephan Müller)
- mgr/dashboard: Use same required field message across the UI ([pr#36277](#), Volker Theile)
- mgr/dashboard: add API team to CODEOWNERS ([pr#36143](#), Ernesto Puerta)
- mgr/dashboard: allow preserving OSD IDs when deleting OSDs ([pr#35766](#), Kiefer Chang)
- mgr/dashboard: cpu stats incorrectly displayed ([pr#36322](#), Avan Thakkar)
- mgr/dashboard: cropped actions menu in nested details ([pr#35620](#), Avan Thakkar)
- mgr/dashboard: fix Source column i18n issue in RBD configuration tables ([pr#35819](#), Kiefer Chang)
- mgr/dashboard: fix backporting issue #35926 ([pr#36073](#), Ernesto Puerta)
- mgr/dashboard: fix pool usage calculation ([pr#36137](#), Ernesto Puerta)
- mgr/dashboard: fix rbdmirroring dropdown menu ([pr#36382](#), Avan Thakkar)
- mgr/dashboard: fix regression in delete OSD modal ([pr#36419](#), Kiefer Chang)
- mgr/dashboard: fix tasks.mgr.dashboard.test\_rbd.RbdTest.test\_move\_image\_to\_trash error ([pr#36563](#), Kiefer Chang)
- mgr/dashboard: fix ui api endpoints ([pr#36160](#), Fabrizio D'Angelo)
- mgr/dashboard: fix wal/db slots controls in the OSD form ([pr#35883](#), Kiefer Chang)
- mgr/dashboard: increase API test coverage in API controllers ([pr#36260](#), Kefu Chai, Aashish Sharma)
- mgr/dashboard: redirect to original URL after successful login ([pr#36831](#), Avan Thakkar)
- mgr/dashboard: remove “This week/month/year” and “Today” time stamps ([pr#36789](#), Avan Thakkar)
- mgr/dashboard: remove cdCopy2ClipboardButton formatted attribute ([pr#35889](#), Tatjana Dehler)
- mgr/dashboard: remove password field if login is using SSO and fix error message in confirm password ([pr#36689](#), Ishan Rai)
- mgr/dashboard: right-align dropdown menu of column filters ([pr#36369](#), Kiefer Chang)
- mgr/dashboard: telemetry activation notification ([pr#35772](#), Tatjana Dehler)

- mgr/dashboard: wait longer for health status to be cleared ([pr#36346](#), Tatjana Dehler)
- mgr/k8sevents: sanitise kubernetes events ([pr#35684](#), Paul Cuzner)
- mgr/prometheus: improve cache ([pr#35847](#), Patrick Seidensal)
- mgr: avoid false alarm of MGR\_MODULE\_ERROR ([pr#35995](#), Kefu Chai)
- mgr: mgr/DaemonServer.cc: make 'config show' on fsid work ([pr#35793](#), Neha Ojha)
- mgr: mgr/cephadm: Adapt Vagrantfile to use octopus instead of master repo on shaman ([pr#35988](#), Volker Theile)
- mgr: mgr/diskprediction\_local: Fix array size error ([pr#36577](#), Benoît Knecht)
- mgr: mgr/progress: Skip pg\_summary update if \_events dict is empty ([pr#36076](#), Manuel Lausch)
- mgr: mgr/prometheus: log time it takes to collect metrics ([pr#36581](#), Patrick Seidensal)
- mgr: mgr: Add missing states to PG\_STATES in mgr\_module.py ([pr#36786](#), Harley Gorrell)
- mgr: mgr: fix race between module load and notify ([pr#35794](#), Mykola Golub)
- mgr: mon/PGMap: do not consider changing pg stuck ([pr#35958](#), Kefu Chai)
- monitoring: alert for pool fill up broken ([pr#35136](#), Volker Theile)
- msgr: New msgr2 crc and secure modes (msgr2.1) ([pr#35720](#), Ilya Dryomov)
- rbd,tests: tests/rbd\_mirror: fix race on test shut down ([pr#36657](#), Mykola Golub)
- rbd: librbd: global and pool-level config overrides require image refresh to apply ([pr#36638](#), Jason Dillaman)
- rbd: librbd: new 'write\_zeroes' API methods to supplement the discard APIs ([pr#36247](#), Jason Dillaman)
- rbd: librbd: potential race conditions handling API IO completions ([pr#36331](#), Jason Dillaman)
- rbd: mgr/dashboard: work with v1 RBD images ([pr#35711](#), Ernesto Puerta)
- rbd: rbd: librbd: Align rbd\_write\_zeroes declarations ([pr#36717](#), Corey Bryant)
- rbd: rbd: librbd: don't resend async\_complete if watcher is unregistered ([pr#36659](#), Mykola Golub)
- rbd: rbd: librbd: flush all queued object IO from simple scheduler ([pr#36658](#),

Jason Dillaman)

- rbd: rbd: librbd: race when disabling object map with overlapping in-flight writes ([pr#36656](#), Jason Dillaman)
- rbd: rbd: recognize crush\_location, read\_from\_replica and compression\_hint map options ([pr#36061](#), Ilya Dryomov)
- rgw,tests: qa/tasks/ragweed: always set ragweed\_repo ([pr#36651](#), Kefu Chai)
- rgw: rgw: lc: fix Segmentation Fault when the tag of the object was not found ([pr#36085](#), yupeng chen, zhuo li)
- rgw: Add subuser to OPA request ([pr#36023](#), Seena Fallah)
- rgw: Add support wildcard subuser for bucket policy ([pr#36022](#), Seena Fallah)
- rgw: Adding data cache and CDN capabilities ([pr#36646](#), Mark Kogan, Or Friedmann)
- rgw: Empty reqs\_change\_state queue before unregistered\_reqs ([pr#36650](#), Soumya Koduri)
- rgw: add abort multipart date and rule-id header to init multipart upload response ([pr#36649](#), zhang Shaowen, zhangshaowen)
- rgw: add access log to the beast frontend ([pr#36024](#), Mark Kogan)
- rgw: add check for index entry's existing when adding bucket stats during bucket reshards ([pr#36025](#), zhang Shaowen)
- rgw: add negative cache to the system object ([pr#36648](#), Or Friedmann)
- rgw: add quota enforcement to CopyObj ([pr#36020](#), Casey Bodley)
- rgw: append obj: prevent tail from being GC'ed ([pr#36389](#), Abhishek Lekshmanan)
- rgw: bucket list/stats truncates for user w/ >1000 buckets ([pr#36019](#), J. Eric Ivancich)
- rgw: cls/rgw: preserve olh entry's name on last unlink ([pr#36652](#), Casey Bodley)
- rgw: cls/rgw\_gc: Fixing the iterator used to access urgent data map ([pr#36017](#), Pritha Srivastava)
- rgw: fix boost::asio::async\_write() does not return error ([pr#36647](#), Mark Kogan)
- rgw: fix bug where ordered bucket listing gets stuck ([pr#35877](#), J. Eric Ivancich)
- rgw: fix double slash (//) killing the gateway ([pr#36654](#), Theofilos Mouratidis)
- rgw: fix loop problem with swift stat on account ([pr#36021](#), Marcus Watts)
- rgw: fix shutdown crash in RGWAsyncReadMDLogEntries ([pr#36653](#), Casey Bodley)

- rgw: introduce safe user-reset-stats ([pr#36655](#), Yuval Lifshitz, Matt Benjamin)
- rgw: lc: add lifecycle perf counters ([pr#36018](#), Mark Kogan, Matt Benjamin)
- rgw: orphan list teuthology test & fully-qualified domain issue ([pr#36027](#), J. Eric Ivancich)
- rgw: orphan-list timestamp fix ([pr#35929](#), J. Eric Ivancich)
- rgw: policy: reuse eval\_principal to evaluate the policy principal ([pr#36636](#), Abhishek Lekshmanan)
- rgw: radoslist incomplete multipart uploads fix marker progression ([pr#36028](#), J. Eric Ivancich)
- rgw: rgw/iam: correcting the result of get role policy ([pr#36645](#), Pritha Srivastava)
- rgw: selinux: allow ceph\_t amqp\_port\_t:tcp\_socket ([pr#36026](#), Kaleb S. KEITHLEY, Thomas Serlin)
- rgw: stop realm reloader before store shutdown ([pr#36644](#), Kefu Chai, Casey Bodley)
- tools: tools: Add statfs operation to ceph-objectstore-tool ([pr#35715](#), David Zafman)

## v15.2.4 Octopus

---

This is the fourth release of the Ceph Octopus stable release series. In addition to a security fix in RGW, this release brings a range of fixes across all components. We recommend that all Octopus users upgrade to this release.

## Notable Changes

---

- CVE-2020-10753: rgw: sanitize newlines in s3 CORSConfiguration's ExposeHeader (William Bowling, Adam Mohammed, Casey Bodley)
- Cephadm: There were a lot of small usability improvements and bug fixes:
  - Grafana when deployed by Cephadm now binds to all network interfaces.
  - `cephadm check-host` now prints all detected problems at once.
  - Cephadm now calls `ceph dashboard set-grafana-api-ssl-verify false` when generating an SSL certificate for Grafana.
  - The Alertmanager is now correctly pointed to the Ceph Dashboard

- `cephadm adopt` now supports adopting an Alertmanager
  - `ceph orch ps` now supports filtering by service name
  - `ceph orch host ls` now marks hosts as offline, if they are not accessible.
- Cephadm can now deploy NFS Ganesha services. For example, to deploy NFS with a service id of `mynfs`, that will use the RADOS pool `nfs-ganesha` and namespace `nfs-ns`:
- ```
1. ceph orch apply nfs mynfs nfs-ganesha nfs-ns
```
- Cephadm: `ceph orch ls --export` now returns all service specifications in yaml representation that is consumable by `ceph orch apply`. In addition, the commands `orch ps` and `orch ls` now support `--format yaml` and `--format json-pretty`.
  - Cephadm: `ceph orch apply osd` supports a `--preview` flag that prints a preview of the OSD specification before deploying OSDs. This makes it possible to verify that the specification is correct, before applying it.
  - RGW: The `radosgw-admin` sub-commands dealing with orphans – `radosgw-admin orphans find`, `radosgw-admin orphans finish`, and `radosgw-admin orphans list-jobs` – have been deprecated. They have not been actively maintained and they store intermediate results on the cluster, which could fill a nearly-full cluster. They have been replaced by a tool, currently considered experimental, `rgw-orphan-list`.
  - RBD: The name of the rbd pool object that is used to store rbd trash purge schedule is changed from “`rbd_trash_trash_purge_schedule`” to “`rbd_trash_purge_schedule`”. Users that have already started using `rbd trash purge schedule` functionality and have per pool or namespace schedules configured should copy “`rbd_trash_trash_purge_schedule`” object to “`rbd_trash_purge_schedule`” before the upgrade and remove “`rbd_trash_purge_schedule`” using the following commands in every RBD pool and namespace where a trash purge schedule was previously configured:
- ```
1. rados -p <pool-name> [-N namespace] cp rbd_trash_trash_purge_schedule rbd_trash_purge_schedule
2. rados -p <pool-name> [-N namespace] rm rbd_trash_trash_purge_schedule
```

or use any other convenient way to restore the schedule after the upgrade.

## Changelog

- build/ops: address SELinux denials observed in rgw/multisite test run ([pr#34538](#), Kefu Chai, Kaleb S. Keithley)
- ceph-volume: add and delete lvm tags in a single lvchange call ([pr#35452](#), Jan Fajerski)

- ceph-volume: add ceph.osdspec\_affinity tag ([pr#35134](#), Joshua Schmid)
- cephadm: batch backport May (1) ([pr#34893](#), Michael Fritch, Ricardo Marques, Matthew Oliver, Sebastian Wagner, Joshua Schmid, Zac Dover, Varsha Rao)
- cephadm: batch backport May (2) ([pr#35188](#), Michael Fritch, Sebastian Wagner, Kefu Chai, Georgios Kyriatsas, Kiefer Chang, Joshua Schmid, Patrick Seidensal, Varsha Rao, Matthew Oliver, Zac Dover, Juan Miguel Olmo Martínez, Tim Serong, Alexey Miasoedov, Ricardo Marques, Satoru Takeuchi)
- cephadm: batch backport June (1) ([pr#35347](#), Sebastian Wagner, Zac Dover, Georgios Kyriatsas, Kiefer Chang, Ricardo Marques, Patrick Seidensal, Patrick Donnelly, Joshua Schmid, Matthew Oliver, Varsha Rao, Juan Miguel Olmo Martínez, Michael Fritch)
- cephadm: batch backport June (2) ([pr#35475](#), Sebastian Wagner, Kiefer Chang, Joshua Schmid, Michael Fritch, shinhwagk, Kefu Chai, Juan Miguel Olmo Martínez, Daniel Pivonka)
- cephfs: allow pool names with hyphen and period ([pr#35251](#), Ramana Raja)
- cephfs: bash\_completion: Do not auto complete obsolete and hidden cmds ([pr#34996](#), Kotresh HR)
- cephfs: cephfs-shell: Change tox testenv name to py3 ([pr#34998](#), Kefu Chai, Varsha Rao, Aditya Srivastava)
- cephfs: client: expose Client::ll\_register\_callback via libcephfs ([pr#35150](#), Jeff Layton)
- cephfs: client: fix Finisher assert failure ([pr#34999](#), Xiubo Li)
- cephfs: client: only set MClientCaps::FLAG\_SYNC when flushing dirty auth caps ([pr#34997](#), Jeff Layton)
- cephfs: fuse: add the '-d' option back for libfuse ([pr#35449](#), Xiubo Li)
- cephfs: mds: Handle blacklisted error in purge queue ([pr#35148](#), Varsha Rao)
- cephfs: mds: preserve ESlaveUpdate logevent until receiving OP\_FINISH ([pr#35253](#), songxinying)
- cephfs: mds: take xlock in the order requests start locking ([pr#35252](#), "Yan, Zheng")
- cephfs: src/client/fuse\_ll: compatible with libfuse3.5 or higher ([pr#35450](#), Jeff Layton, Xiubo Li)
- cephfs: vstart\_runner: set mounted to True at the end of mount() ([pr#35447](#), Rishabh Dave)

- core: bluestore: fix large (>2GB) writes when bluefs\_buffered\_io = true ([pr#35446](#), Igor Fedotov)
- core: bluestore: introduce hybrid allocator ([pr#35498](#), Igor Fedotov, Adam Kupczyk)
- core: cls/queue: fix empty markers when listing entries ([pr#35241](#), Pritha Srivastava, Yuval Lifshitz)
- core: objecter: don't attempt to read from non-primary on EC pools ([pr#35444](#), Ilya Dryomov)
- core: osd: add -osdspec-affinity flag ([pr#35382](#), Joshua Schmid)
- core: osd: make "missing incremental map" a debug log message ([pr#35442](#), Nathan Cutler)
- core: osd: prevent ShardedOpWQ suicide\_grace drop when waiting for work ([pr#34881](#), Dan Hill)
- core: rocksdb: Update to ceph-octopus-v5.8-1436 ([pr#35036](#), Brad Hubbard)
- doc: drop obsolete cache tier options ([pr#35105](#), Nathan Cutler)
- doc: mgr/dashboard: Add troubleshooting guide ([pr#34947](#), Tatjana Dehler)
- doc: rgw: document 'rgw gc max concurrent io' ([pr#34987](#), Casey Bodley)
- mds: cleanup uncommitted fragments before mds goes to active ([pr#35448](#), "Yan, Zheng")
- mds: don't assert empty io context list when shutting down ([pr#34509](#), "Yan, Zheng")
- mds: don't shallow copy when decoding xattr map ([pr#35147](#), "Yan, Zheng")
- mds: flag backtrace scrub failures for new files as okay ([pr#35555](#), Milind Changire)
- mgr/dashboard/grafana: Add rbd-image details dashboard ([pr#35247](#), Enno Gotthold)
- mgr/dashboard: Asynchronous unique username validation for User Component ([pr#34849](#), Nizamudeen)
- mgr/dashboard: ECP modal enhancement ([pr#35152](#), Stephan Müller)
- mgr/dashboard: Fix HomeTest setup ([pr#35085](#), Tiago Melo)
- mgr/dashboard: Fix e2e chromium binary validation ([pr#35679](#), Tiago Melo)
- mgr/dashboard: Fix random E2E error in mgr-modules ([pr#35706](#), Tiago Melo)

- mgr/dashboard: Fix redirect after changing password ([pr#35243](#), Tiago Melo)
- mgr/dashboard: Prevent dashboard breakdown on bad pool selection ([pr#35135](#), Stephan Müller)
- mgr/dashboard: Proposed About Modal box ([pr#35291](#), Ngwa Sedrick Meh, Tiago Melo)
- mgr/dashboard: Reduce requests in Mirroring page ([pr#34992](#), Tiago Melo)
- mgr/dashboard: Replace Protractor with Cypress ([pr#34910](#), Tiago Melo)
- mgr/dashboard: Show labels in hosts page ([pr#35517](#), Volker Theile)
- mgr/dashboard: Show table details inside the datatable ([pr#35270](#), Sebastian Krah)
- mgr/dashboard: add telemetry report component ([pr#34850](#), Tatjana Dehler)
- mgr/dashboard: displaying Service detail inside table ([pr#35269](#), Kiefer Chang)
- mgr/dashboard: fix autocomplete input backgrounds in chrome and firefox ([pr#35718](#), Ishan Rai)
- mgr/dashboard: grafana panels for rgw multisite sync performance ([pr#35693](#), Alfonso Martínez)
- mgr/dashboard: monitoring menu entry should indicate firing alerts ([pr#34822](#), Tiago Melo, Volker Theile)
- mgr/dashboard: redesign the login screen ([pr#35268](#), Ishan Rai)
- mgr/dashboard: remove space after service name in the Hosts List table ([pr#35531](#), Kiefer Chang)
- mgr/dashboard: replace hard coded telemetry URLs ([pr#35231](#), Tatjana Dehler)
- mgr/rbd\_support: rename “rbd\_trash\_trash\_purge\_schedule” oid ([pr#35436](#), Nathan Cutler, Mykola Golub)
- mgr/status: Fix “ceph fs status” json format writing to stderr ([pr#34727](#), Kotresh HR)
- mgr/test\_orchestrator: fix \_get\_ceph\_daemons() ([pr#34979](#), Alfonso Martínez)
- mgr/volumes: Add snapshot info command ([pr#35670](#), Kotresh HR)
- mgr/volumes: Create subvolume with isolated rados namespace ([pr#35671](#), Kotresh HR)
- mgr/volumes: Fix subvolume create idempotency ([pr#35256](#), Kotresh HR)
- mgr: synchronize ClusterState’s health and mon\_status ([pr#34995](#), Radosław Zarzynski)

- monitoring: Fix “10% OSDs down” alert description ([pr#35151](#), Benoît Knecht)
- monitoring: fixing some issues in RBD detail dashboard ([pr#35463](#), Kiefer Chang)
- rbd: librbd: Watcher should not attempt to re-watch after detecting blacklisting ([pr#35439](#), Jason Dillaman)
- rbd: librbd: avoid completing mirror:DisableRequest while holding its lock ([pr#35126](#), Jason Dillaman)
- rbd: librbd: copy API should not inherit v1 image format by default ([pr#35255](#), Jason Dillaman)
- rbd: librbd: make rbd\_read\_from\_replica\_policy actually work ([pr#35438](#), Ilya Dryomov)
- rbd: pybind: RBD.create() method’s ‘old\_format’ parameter now defaults to False ([pr#35435](#), Jason Dillaman)
- rbd: rbd-mirror: don’t hold (stale) copy of local image journal pointer ([pr#35430](#), Jason Dillaman)
- rbd: rbd-mirror: stop local journal replayer first during shut down ([pr#35440](#), Jason Dillaman, Mykola Golub)
- rbd: rbd-mirror: wait for in-flight start/stop/restart ([pr#35437](#), Mykola Golub)
- rgw: add “rgw-orphan-list” tool and “radosgw-admin bucket radoslist ...” ([pr#34991](#), J. Eric Ivancich)
- rgw: amqp: fix the “routable” delivery mode ([pr#35433](#), Yuval Lifshitz)
- rgw: anonymous swift to obj that dont exist should 401 ([pr#35120](#), Matthew Oliver)
- rgw: fix bug where bucket listing end marker not always set correctly ([pr#34993](#), J. Eric Ivancich)
- rgw: fix rgw tries to fetch anonymous user ([pr#34988](#), Or Friedmann)
- rgw: fix some list buckets handle leak ([pr#34985](#), Tianshan Qu)
- rgw: gc: Clearing off urgent data in bufferlist, before ([pr#35434](#), Pritha Srivastava)
- rgw: lc: enable thread-parallelism in RGWLC ([pr#35431](#), Matt Benjamin)
- rgw: notifications: fix zero size in notifications ([pr#34940](#), J. Eric Ivancich, Yuval Lifshitz)
- rgw: notifications: version id was not sent in versioned buckets ([pr#35254](#), Yuval Lifshitz)

- rgw: radosgw-admin: fix infinite loops in ‘datalog list’ ([pr#34989](#), Casey Bodley)
- rgw: url: fix amqp urls with vhosts ([pr#35432](#), Yuval Lifshitz)
- tests: migrate qa/ to Python3 ([pr#35364](#), Kyr Shatskyy, Ilya Dryomov, Xiubo Li, Kefu Chai, Casey Bodley, Rishabh Dave, Patrick Donnelly, Sidharth Anupkrishnan, Michael Fritch)

## v15.2.3 Octopus

---

This is the third bug-fix release of the Ceph Octopus stable release series. This release mainly is a workaround for a potential OSD corruption in v15.2.2. We advise users to upgrade to v15.2.3 directly. For users running v15.2.2 please execute the following:

```
1. ceph config set osd bluefs_preextent_wal_files false
```

## Changelog

---

- bluestore: remove preextended WAL support ([issue#45613](#), Igor Fedotov, Neha Ojha)

## v15.2.2 Octopus

---

This is the second bug-fix release of the Ceph Octopus stable release series. This release brings a range of fixes across all components, as well as patching a security flaw. We recommend that all Octopus users upgrade.

## Notable Changes

---

- CVE-2020-10736: Fixed an authorization bypass in mons & mgrs (Olle SegerDahl, Josh Durgin)

## Changelog

---

- bluestore,core: common/options: Disable bluefs\_buffered\_io by default again ([pr#34353](#), Mark Nelson)
- bluestore: os/bluestore: Don’t pollute old journal when add new device ([pr#34795](#), Yang Honggang)
- bluestore: os/bluestore: fix ‘unused’ calculation ([pr#34793](#), Igor Fedotov, xie xingguo)
- bluestore: os/bluestore: open DB in read-only when expanding DB/WAL ([pr#34610](#),

Adam Kupczyk, Igor Fedotov)

- build/ops: rpm: add python3-saml as install dependency ([pr#34474](#), Ernesto Puerta)
- build/ops: rpm: drop “is\_opensuse” conditional in SUSE-specific bcond block ([pr#34790](#), Nathan Cutler)
- build/ops: spec: address some warnings raised by RPM 4.15.1 ([pr#34526](#), Nathan Cutler)
- ceph-volume/batch: check lvs list before access ([pr#34480](#), Jan Fajerski)
- ceph-volume/batch: return success when all devices are filtered ([pr#34477](#), Jan Fajerski)
- ceph-volume: update functional testing deploy.yml playbook ([pr#34886](#), Guillaume Abrioux)
- cephadm: Fix check\_ip\_port to work with IPv6 ([pr#34350](#), Ricardo Marques)
- cephadm: Update images used ([pr#34686](#), Sebastian Wagner)
- cephadm: ceph-volume: disallow concurrent execution ([pr#34423](#), Sage Weil)
- cephadm: rm-cluster clean up /etc/ceph ([pr#34299](#), Daniel-Pivonka)
- cephfs,mgr: mgr/volumes: Add interface to get subvolume metadata ([pr#34681](#), Kotresh HR)
- cephfs,mgr: mgr: force purge normal ceph entities from service map ([issue#44677](#), [pr#34800](#), Venky Shankar)
- cephfs,tools: cephfs-journal-tool: correctly parse -dry\_run argument ([pr#34804](#), Milind Changire)
- cephfs,tools: tools/cephfs: add accounted\_rstat/rstat when building file dentry ([pr#34803](#), Xiubo Li)
- cephfs: ceph-fuse: link to libfuse3 and pass -o big\_writes to libfuse if libfuse < 3.0.0 ([pr#34769](#), Xiubo Li, “Yan, Zheng”, Kefu Chai)
- cephfs: client: reset requested\_max\_size if file write is not wanted ([pr#34766](#), “Yan, Zheng”)
- cephfs: mds: fix ‘if there is lock cache on dir’ check ([pr#34273](#), “Yan, Zheng”)
- cephfs: mon/FSCommands: Fix ‘add\_data\_pool’ command and ‘fs new’ command ([pr#34775](#), Ramana Raja)
- cephfs: qa: install task runs twice with double unwind causing fatal errors ([pr#34912](#), Patrick Donnelly)

- core,mon: mon/OSDMonitor: allow trimming maps even if osds are down ([pr#34924](#), Joao Eduardo Luis)
- core: ceph-object-corpus: update to octopus ([pr#34797](#), Josh Durgin)
- core: mgr/DaemonServer: fetch metadata for new daemons (e.g., mons) ([pr#34416](#), Sage Weil)
- core: mon/OSDMonitor: Always tune priority cache manager memory on all mons ([pr#34917](#), Sridhar Seshasayee)
- core: mon: calculate min\_size on osd pool set size ([pr#34528](#), Deepika Upadhyay)
- core: osd/PeeringState: do not trim pg log past last\_update\_ondisk ([pr#34807](#), xie xingguo, Samuel Just)
- core: osd/PrimaryLogPG: fix SPARSE\_READ stat ([pr#34809](#), Yan Jun)
- devices/simple/scan: Fix string in log statement ([pr#34446](#), Jan Fajerski)
- doc: cephadm: Batch backport April (1) ([pr#34554](#), Matthew Oliver, Sage Weil, Sebastian Wagner, Michael Fritch, Tim, Jeff Layton, Juan Miguel Olmo Martínez, Joshua Schmid)
- doc: cephadm: Batch backport April (2) ([issue#45029](#), [pr#34687](#), Maran Hidskes, Kiefer Chang, Matthew Oliver, Sebastian Wagner, Andreas Haase, Tim Serong, Zac Dover, Michael Fritch, Joshua Schmid)
- doc: cephadm: Batch backport April (3) ([pr#34742](#), Sebastian Wagner, Dimitri Savineau, Michael Fritch)
- doc: cephadm: batch backport March ([pr#34438](#), Jan Fajerski, Sebastian Wagner, Daniel-Pivonka, Michael Fritch, Sage Weil)
- doc: doc/releases/nutilus: restart OSDs to make them bind to v2 addr ([pr#34523](#), Nathan Cutler)
- mgr/dashboard: 'Prometheus / All Alerts' page shows progress bar ([pr#34631](#), Volker Theile)
- mgr/dashboard: Fix ServiceDetails and PoolDetails unit tests ([pr#34760](#), Tiago Melo)
- mgr/dashboard: Fix iSCSI's username and password validation ([pr#34547](#), Tiago Melo)
- mgr/dashboard: Improve iSCSI CHAP message ([pr#34630](#), Ricardo Marques)
- mgr/dashboard: Prevent iSCSI target recreation when editing controls ([pr#34548](#), Tiago Melo)
- mgr/dashboard: RGW auto refresh is not working ([pr#34739](#), Avan Thakkar)

- mgr/dashboard: Repair broken grafana panels ([pr#34495](#), Kristoffer Grönlund)
- mgr/dashboard: Update translations on octopus ([pr#34309](#), Sebastian Krah)
- mgr/dashboard: add crush rule test suite ([pr#34211](#), Tatjana Dehler)
- mgr/dashboard: fix API tests to be py3 compatible ([pr#34759](#), Kefu Chai, Laura Paduano, Alfonso Martínez)
- mgr/dashboard: fix errors related to frontend service subscriptions ([pr#34467](#), Alfonso Martínez)
- mgr/dashboard: fix tasks.mgr.dashboard.test\_rgw.RgwBucketTest.test\_all ([pr#34708](#), Alfonso Martínez)
- mgr/dashboard: lint error on plugins/debug.py ([pr#34625](#), Volker Theile)
- mgr/dashboard: shorten “Container ID” and “Container image ID” in Services page ([pr#34648](#), Volker Theile)
- mgr/dashboard: use FQDN for failover redirection ([pr#34498](#), Ernesto Puerta)
- mgr: mgr/PyModule: fix missing tracebacks in handle\_pyerror() ([pr#34626](#), Tim Serong)
- mgr: mgr/telegraf: catch FileNotFoundError exception ([pr#34629](#), Kefu Chai)
- monitoring: Fix pool capacity incorrect ([pr#34449](#), James Cheng)
- monitoring: alert for prediction of disk and pool fill up broken ([pr#34395](#), Patrick Seidensal)
- monitoring: fix decimal precision in Grafana %percentages ([pr#34828](#), Ernesto Puerta)
- monitoring: root volume full alert fires false positives ([pr#34418](#), Patrick Seidensal)
- pybind,rbd: pybind/rbd: ensure image is open before permitting operations ([pr#34425](#), Mykola Golub)
- pybind,rbd: pybind/rbd: fix no lockers are obtained, ImageNotFound exception will be output ([pr#34387](#), zhangdaolong)
- qa/suites/rados/cephadm/upgrade: start from v15.2.0 ([pr#34440](#), Sage Weil)
- qa/tasks/cephadm: add ‘roleless’ mode ([pr#34407](#), Sage Weil)
- rbd,tests: tests: update unmap.t for table spacing changes ([pr#34819](#), Ilya Dryomov)
- rbd: rbd-mirror: improved replication statistics ([pr#34810](#), Mykola Golub, Jason

Dillaman)

- rbd: rbd: ignore tx-only mirror peers when adding new peers ([pr#34638](#), Jason Dillaman)
- rgw: Disable prefetch of entire head object when GET request with range header ([pr#34826](#), Or Friedmann)
- rgw: pubsub sync module ignores ERR\_USER\_EXIST ([pr#34825](#), Casey Bodley)
- rgw: radosgw-admin: add support for -bucket-id in bucket stats command ([pr#34816](#), Vikhyat Umrao)
- rgw: reshards: skip stale bucket id entries from reshards queue ([pr#34734](#), Abhishek Lekshmanan)
- rgw: use DEFER\_DROP\_PRIVILEGES flag unconditionally ([pr#34731](#), Casey Bodley)

## v15.2.1 Octopus

This is the first bugfix release of Ceph Octopus, we recommend all Octopus users upgrade. This release fixes an upgrade issue and also has 2 security fixes

## Notable Changes

- issue#44759: Fixed luminous->nautilus->octopus upgrade asserts
- CVE-2020-1759: Fixed nonce reuse in msgr V2 secure mode
- CVE-2020-1760: Fixed XSS due to RGW GetObject header-splitting

## Changelog

- build/ops: fix ceph\_release type to 'stable' ([pr#34194](#), Sage Weil)
- build/ops: vstart\_runner.py: fix OSError when checking if non-existent path is mounted ([pr#34132](#), Alfonso Martínez)
- cephadm: Add alertmanager adopt ([pr#34157](#), Eric Jackson)
- cephadm: Add alertmanager sample ([pr#34158](#), Eric Jackson)
- cephadm: Fix truncated output of "ceph mgr dump" ([pr#34258](#), Sebastian Wagner)
- mgr/cephadm: Add example to run when debugging ssh failures ([pr#34153](#), Sebastian Wagner)
- mgr/cephadm: DriveGroupSpec needs to support/ignore \_unmanaged\_ ([pr#34185](#), Joshua Schmid)
- mgr/cephadm: bind grafana to all interfaces ([pr#34191](#), Sage Weil)
- mgr/cephadm: fix 'orch ps -refresh' ([pr#34190](#), Sage Weil)
- mgr/cephadm: fix 'upgrade start' message when specifying a version ([pr#34186](#), Sage Weil)
- mgr/cephadm: include alerts in prometheus deployment ([pr#34155](#), Sage Weil)
- mgr/cephadm: point alertmanager at all mgr/dashboard URLs ([pr#34154](#), Sage Weil)
- mgr/cephadm: provision nfs-ganesha via orchestrator ([pr#34192](#), Michael Fritch)
- mgr/dashboard: Check for missing npm resolutions ([pr#34202](#), Tiago Melo)
- mgr/dashboard: NoRebalance flag is added to the Dashboard ([pr#33939](#), Nizamudeen)

- mgr/dashboard: correct Orchestrator documentation link ([pr#34212](#), Tatjana Dehler)
- mgr/dashboard: do not fail on user creation (CLI) ([pr#34280](#), Tatjana Dehler)
- mgr/orch: allow list daemons by service\_name ([pr#34160](#), Kiefer Chang)
- mgr/prometheus: ceph\_pg\_\* metrics contains last value instead of sum across all reported states ([pr#34163](#), Jacek Suchenia)
- mgr/rook: Blinking lights ([pr#34199](#), Juan Miguel Olmo Martinez)
- osd/PeeringState: drop mimic assert ([pr#34204](#), Sage Weil)
- osd/PeeringState: fix pending want\_acting vs osd offline race ([pr#34123](#), xie xingguo)
- pybind/mgr: fix config\_notify handling of default values ([pr#34178](#), Nathan Cutler)
- rbd: librbd: fix client backwards compatibility issues ([issue#39450](#), [issue#38834](#), [pr#34323](#), Jason Dillaman)
- tools: ceph-backport.sh: add deprecation warning ([pr#34125](#), Nathan Cutler)

## v15.2.0 Octopus

---

This is the first stable release of Ceph Octopus.

## Major Changes from Nautilus

---

### General

- A new deployment tool called **cephadm** has been introduced that integrates Ceph daemon deployment and management via containers into the orchestration layer. For more information see [Cephadm](#).
- Health alerts can now be muted, either temporarily or permanently.
- Health alerts are now raised for recent Ceph daemons crashes.
- A simple ‘alerts’ module has been introduced to send email health alerts for clusters deployed without the benefit of an existing external monitoring infrastructure.
- [Packages](#) are built for the following distributions:
  - CentOS 8
  - CentOS 7 (partial-see below)

- Ubuntu 18.04 (Bionic)
- Debian Buster
- Container image (based on CentOS 8)

Note that the dashboard, prometheus, and restful manager modules will not work on the CentOS 7 build due to Python 3 module dependencies that are missing in CentOS 7.

Besides this packages built by the community will also available for the following distros:

- Fedora (33/rawhide)
- openSUSE (15.2, Tumbleweed)

## Dashboard

The [Ceph Dashboard](#) has gained a lot of new features and functionality:

- UI Enhancements
  - New vertical navigation bar
  - New unified sidebar: better background task and events notification
  - Shows all progress mgr module notifications
  - Multi-select on tables to perform bulk operations
- Dashboard user account security enhancements
  - Disabling/enabling existing user accounts
  - Clone an existing user role
  - Users can change their own password
  - Configurable password policies: Minimum password complexity/length requirements
  - Configurable password expiration
  - Change password after first login

New and enhanced management of Ceph features/services:

- OSD/device management
  - List all disks associated with an OSD
  - Add support for blinking enclosure LEDs via the orchestrator

- List all hosts known by the orchestrator
  - List all disks and their properties attached to a node
  - Display disk health information (health prediction and SMART data)
  - Deploy new OSDs on new disks/hosts
  - Display and allow sorting by an OSD's default device class in the OSD table
  - Explicitly set/change the device class of an OSD, display and sort OSDs by device class
- Pool management
    - Viewing and setting pool quotas
    - Define and change per-pool PG autoscaling mode
  - RGW management enhancements
    - Enable bucket versioning
    - Enable MFA support
    - Select placement target on bucket creation
  - CephFS management enhancements
    - CephFS client eviction
    - CephFS snapshot management
    - CephFS quota management
    - Browse CephFS directory
  - iSCSI management enhancements
    - Show iSCSI GW status on landing page
    - Prevent deletion of IQNs with open sessions
    - Display iSCSI "logged in" info
  - Prometheus alert management
    - List configured Prometheus alerts

## RADOS

- Objects can now be brought in sync during recovery by copying only the modified portion of the object, reducing tail latencies during recovery.

- Ceph will allow recovery below *min\_size* for Erasure coded pools, wherever possible.
- The PG autoscaler feature introduced in Nautilus is enabled for new pools by default, allowing new clusters to autotune *pg num* without any user intervention. The default values for new pools and RGW/CephFS metadata pools have also been adjusted to perform well for most users.
- BlueStore has received several improvements and performance updates, including improved accounting for “omap” (key/value) object data by pool, improved cache memory management, and a reduced allocation unit size for SSD devices. (Note that by default, the first time each OSD starts after upgrading to octopus it will trigger a conversion that may take from a few minutes to a few hours, depending on the amount of stored “omap” data.)
- Snapshot trimming metadata is now managed in a more efficient and scalable fashion.

## RBD block storage

- Mirroring now supports a new snapshot-based mode that no longer requires the journaling feature and its related impacts in exchange for the loss of point-in-time consistency (it remains crash consistent).
- Clone operations now preserve the sparseness of the underlying RBD image.
- The trash feature has been improved to (optionally) automatically move old parent images to the trash when their children are all deleted or flattened.
- The trash can be configured to automatically purge on a defined schedule.
- Images can be online re-sparsified to reduce the usage of zeroed extents.
- The `rbd-nbd` tool has been improved to use more modern kernel interfaces.
- Caching has been improved to be more efficient and performant.
- `rbd-mirror` automatically adjusts its per-image memory usage based upon its memory target.
- A new persistent read-only caching daemon is available to offload reads from shared parent images.

## RGW object storage

- New [Multisite Sync Policy](#) primitives for per-bucket replication. (EXPERIMENTAL)
- S3 feature support:
  - Bucket Replication (EXPERIMENTAL)
  - [Bucket Notifications](#) via HTTP/S, AMQP and Kafka
  - Bucket Tagging
  - Object Lock
  - Public Access Block for buckets
- Bucket sharding:
  - Significantly improved listing performance on buckets with many shards.
  - Dynamic resharding prefers prime shard counts for improved distribution.
  - Raised the default number of bucket shards to 11.
- Added [HashiCorp Vault Integration](#) for SSE-KMS.

- Added Keystone token cache for S3 requests.

## CephFS distributed file system

- Inline data support in CephFS has been deprecated and will likely be removed in a future release.
- MDS daemons can now be assigned to manage a particular file system via the new `mds_join_fs` option.
- MDS now aggressively asks idle clients to trim caps which improves stability when file system load changes.
- The mgr volumes plugin has received numerous improvements to support CephFS via CSI, including snapshots and cloning.
- cephfs-shell has had numerous incremental improvements and bug fixes.

## Upgrading from Mimic or Nautilus

---

### Note

You can monitor the progress of your upgrade at each stage with the `ceph versions` command, which will tell you what ceph version(s) are running for each type of daemon.

# Instructions

1. Make sure your cluster is stable and healthy (no down or recovering OSDs). (Optional, but recommended.)
2. Set the `noout` flag for the duration of the upgrade. (Optional, but recommended.):

```
1. # ceph osd set noout
```

3. Upgrade monitors by installing the new packages and restarting the monitor daemons. For example, on each monitor host,:

```
1. # systemctl restart ceph-mon.target
```

Once all monitors are up, verify that the monitor upgrade is complete by looking for the `octopus` string in the mon map. The command:

```
1. # ceph mon dump | grep min_mon_release
```

should report:

```
1. min_mon_release 15 (octopus)
```

If it doesn't, that implies that one or more monitors hasn't been upgraded and restarted and/or the quorum does not include all monitors.

4. Upgrade `ceph-mgr` daemons by installing the new packages and restarting all manager daemons. For example, on each manager host,:

```
1. # systemctl restart ceph-mgr.target
```

Verify the `ceph-mgr` daemons are running by checking `ceph -s`:

```
1. # ceph -s
2.
3. ...
4. services:
5.   mon: 3 daemons, quorum foo,bar,baz
6.   mgr: foo(active), standbys: bar, baz
7. ...
```

5. Upgrade all OSDs by installing the new packages and restarting the `ceph-osd` daemons on all OSD hosts:

```
1. # systemctl restart ceph-osd.target
```

Note that the first time each OSD starts, it will do a format conversion to improve the accounting for “omap” data. This may take a few minutes to as much as a few hours (for an HDD with lots of omap data). You can disable this automatic conversion with:

```
1. # ceph config set osd bluestore_fsck_quick_fix_on_mount false
```

You can monitor the progress of the OSD upgrades with the [ceph versions](#) or [ceph osd versions](#) commands:

```
1. # ceph osd versions
2. {
3.     "ceph version 13.2.5 (...) mimic (stable)": 12,
4.     "ceph version 15.2.0 (...) octopus (stable)": 22,
5. }
```

6. Upgrade all CephFS MDS daemons. For each CephFS file system,

i. Reduce the number of ranks to 1. (Make note of the original number of MDS daemons first if you plan to restore it later.):

```
1. # ceph status
2. # ceph fs set <fs_name> max_mds 1
```

ii. Wait for the cluster to deactivate any non-zero ranks by periodically checking the status:

```
1. # ceph status
```

iii. Take all standby MDS daemons offline on the appropriate hosts with:

```
1. # systemctl stop ceph-mds@<daemon_name>
```

iv. Confirm that only one MDS is online and is rank 0 for your FS:

```
1. # ceph status
```

v. Upgrade the last remaining MDS daemon by installing the new packages and restarting the daemon:

```
1. # systemctl restart ceph-mds.target
```

vi. Restart all standby MDS daemons that were taken offline:

```
1. # systemctl start ceph-mds.target
```

vii. Restore the original value of `max_mds` for the volume:

```
1. # ceph fs set <fs_name> max_mds <original_max_mds>
```

7. Upgrade all radosgw daemons by upgrading packages and restarting daemons on all hosts:

```
1. # systemctl restart ceph-radosgw.target
```

8. Complete the upgrade by disallowing pre-Octopus OSDs and enabling all new Octopus-only functionality:

```
1. # ceph osd require-osd-release octopus
```

9. If you set `noout` at the beginning, be sure to clear it with:

```
1. # ceph osd unset noout
```

10. Verify the cluster is healthy with `ceph health`.

If your CRUSH tunables are older than Hammer, Ceph will now issue a health warning. If you see a health alert to that effect, you can revert this change with:

```
1. ceph config set mon mon_crush_min_required_version firefly
```

If Ceph does not complain, however, then we recommend you also switch any existing CRUSH buckets to straw2, which was added back in the Hammer release. If you have any 'straw' buckets, this will result in a modest amount of data movement, but generally nothing too severe.:

```
1. ceph osd getcrushmap -o backup-crushmap
2. ceph osd crush set-all-straw-buckets-to-straw2
```

If there are problems, you can easily revert with:

```
1. ceph osd setcrushmap -i backup-crushmap
```

Moving to 'straw2' buckets will unlock a few recent features, like the crush-compat `balancer` mode added back in Luminous.

11. If you are upgrading from Mimic, or did not already do so when you upgraded to Nautlius, we recommended you enable the new `v2 network protocol`, issue the

following command:

```
1. ceph mon enable-msgr2
```

This will instruct all monitors that bind to the old default port 6789 for the legacy v1 protocol to also bind to the new 3300 v2 protocol port. To see if all monitors have been updated,:

```
1. ceph mon dump
```

and verify that each monitor has both a `v2:` and `v1:` address listed.

12. Consider enabling the [telemetry module](#) to send anonymized usage statistics and crash information to the Ceph upstream developers. To see what would be reported (without actually sending any information to anyone),:

```
1. ceph mgr module enable telemetry  
2. ceph telemetry show
```

If you are comfortable with the data that is reported, you can opt-in to automatically report the high-level cluster metadata with:

```
1. ceph telemetry on
```

For more information about the telemetry module, see [the documentation](#).

## Upgrading from pre-Mimic releases (like Luminous)

You *must* first upgrade to Mimic (13.2.z) or Nautilus (14.2.z) before upgrading to Octopus.

# Upgrade compatibility notes

- Starting with Octopus, there is now a separate repository directory for each version on download.ceph.com (e.g., `rpm-15.2.0` and `debian-15.2.0`). The traditional package directory that is named after the release (e.g., `rpm-octopus` and `debian-octopus`) is now a symlink to the most recently bug fix version for that release. We no longer generate a single repository that combines all bug fix versions for a single named release.
- The RGW “num\_rados\_handles” has been removed. If you were using a value of “num\_rados\_handles” greater than 1 multiply your current “objecter\_inflight\_ops” and “objecter\_inflight\_op\_bytes” parameters by the old “num\_rados\_handles” to get the same throttle behavior.
- Ceph now packages python bindings for python3.6 instead of python3.4, because python3 in EL7/EL8 is now using python3.6 as the native python3. see the [announcement](#) for more details on the background of this change.
- librbd now uses a write-around cache policy by default, replacing the previous write-back cache policy default. This cache policy allows librbd to immediately complete write IOs while they are still in-flight to the OSDs. Subsequent flush requests will ensure all in-flight write IOs are completed prior to completing. The librbd cache policy can be controlled via a new “rbd\_cache\_policy” configuration option.
- librbd now includes a simple IO scheduler which attempts to batch together multiple IOs against the same backing RBD data block object. The librbd IO scheduler policy can be controlled via a new “rbd\_io\_scheduler” configuration option.
- RGW: radosgw-admin introduces two subcommands that allow the managing of expire-stale objects that might be left behind after a bucket reshards in earlier versions of RGW. One subcommand lists such objects and the other deletes them. Read the troubleshooting section of the dynamic resharding docs for details.
- RGW: Bucket naming restrictions have changed and likely to cause `InvalidBucketName` errors. We recommend to set `rgw_relaxed_s3_bucket_names` option to true as a workaround.
- In the Zabbix Mgr Module there was a typo in the key being sent to Zabbix for PGs in `backfill_wait` state. The key that was sent was ‘`wait_backfill`’ and the correct name is ‘`backfill_wait`’. Update your Zabbix template accordingly so that it accepts the new key being sent to Zabbix.
- zabbix plugin for ceph manager now includes osd and pool discovery. Update of `zabbix_template.xml` is needed to receive per-pool (read/write throughput, diskspace usage) and per-osd (latency, status, pgs) statistics

- The format of all date + time stamps has been modified to fully conform to ISO 8601. The old format (`YYYY-MM-DD HH:MM:SS.ssssss`) excluded the `T` separator between the date and time and was rendered using the local time zone without any explicit indication. The new format includes the separator as well as a `+nnnn` or `-nnnn` suffix to indicate the time zone, or a `Z` suffix if the time is UTC. For example, `2019-04-26T18:40:06.225953+0100`.

Any code or scripts that was previously parsing date and/or time values from the JSON or XML structure CLI output should be checked to ensure it can handle ISO 8601 conformant values. Any code parsing date or time values from the unstructured human-readable output should be modified to parse the structured output instead, as the human-readable output may change without notice.

- The `bluestore_no_per_pool_stats_tolerance` config option has been replaced with `bluestore_fsck_error_on_no_per_pool_stats` (default: false). The overall default behavior has not changed: fsck will warn but not fail on legacy stores, and repair will convert to per-pool stats.
- The disaster-recovery related ‘ceph mon sync force’ command has been replaced with ‘ceph daemon <...> sync\_force’.
- The `osd_recovery_max_active` option now has `osd_recovery_max_active_hdd` and `osd_recovery_max_active_ssd` variants, each with different default values for HDD and SSD-backed OSDs, respectively. By default `osd_recovery_max_active` now defaults to zero, which means that the OSD will conditionally use the HDD or SSD option values. Administrators who have customized this value may want to consider whether they have set this to a value similar to the new defaults (3 for HDDs and 10 for SSDs) and, if so, remove the option from their configuration entirely.
- monitors now have a `ceph osd info` command that will provide information on all osds, or provided osds, thus simplifying the process of having to parse `osd dump` for the same information.
- The structured output of `ceph status` or `ceph -s` is now more concise, particularly the mgrmap and monmap sections, and the structure of the osdmap section has been cleaned up.
- A health warning is now generated if the average osd heartbeat ping time exceeds a configurable threshold for any of the intervals computed. The OSD computes 1 minute, 5 minute and 15 minute intervals with average, minimum and maximum values. New configuration option `mon_warn_on_slow_ping_ratio` specifies a percentage of `osd_heartbeat_grace` to determine the threshold. A value of zero disables the warning. New configuration option `mon_warn_on_slow_ping_time` specified in milliseconds over-rides the computed value, causes a warning when OSD heartbeat pings take longer than the specified amount. New admin command `ceph daemon mgr.# dump_osd_network [threshold]` command will list all connections with a ping time longer than the specified threshold or value determined by the config options, for the average for any of the 3 intervals. New admin command `ceph daemon osd.# dump_osd_network [threshold]`

will do the same but only including heartbeats initiated by the specified OSD.

- Inline data support for CephFS has been deprecated. When setting the flag, users will see a warning to that effect, and enabling it now requires the `--yes-i-really-mean-it` flag. If the MDS is started on a filesystem that has it enabled, a health warning is generated. Support for this feature will be removed in a future release.
- `ceph {set unset} full` is not supported anymore. We have been using `full` and `nearfull` flags in OSD map for tracking the fullness status of a cluster back since the Hammer release, if the OSD map is marked `full` all write operations will be blocked until this flag is removed. In the Infernalis release and Linux kernel 4.7 client, we introduced the per-pool full/nearfull flags to track the status for a finer-grained control, so the clients will hold the write operations if either the cluster-wide `full` flag or the per-pool `full` flag is set. This was a compromise, as we needed to support the cluster with and without per-pool `full` flags support. But this practically defeated the purpose of introducing the per-pool flags. So, in the Mimic release, the new flags finally took the place of their cluster-wide counterparts, as the monitor started removing these two flags from OSD map. So the clients of Infernalis and up can benefit from this change, as they won't be blocked by the full pools which they are not writing to. In this release, `ceph {set unset} full` is now considered as an invalid command. And the clients will continue honoring both the cluster-wide and per-pool flags to be backward compatible with pre-infernalis clusters.
- The telemetry module now reports more information.

First, there is a new ‘device’ channel, enabled by default, that will report anonymized hard disk and SSD health metrics to `telemetry.ceph.com` in order to build and improve device failure prediction algorithms. If you are not comfortable sharing device metrics, you can disable that channel first before re-opting-in:

```
1. ceph config set mgr mgr/telemetry/channel_device false
```

Second, we now report more information about CephFS file systems, including:

- how many MDS daemons (in total and per file system)
- which features are (or have been) enabled
- how many data pools
- approximate file system age (year + month of creation)
- how many files, bytes, and snapshots
- how much metadata is being cached

We have also added:

- which Ceph release the monitors are running
- whether msgr v1 or v2 addresses are used for the monitors
- whether IPv4 or IPv6 addresses are used for the monitors
- whether RADOS cache tiering is enabled (and which mode)
- whether pools are replicated or erasure coded, and which erasure code profile plugin and parameters are in use
- how many hosts are in the cluster, and how many hosts have each type of daemon
- whether a separate OSD cluster network is being used
- how many RBD pools and images are in the cluster, and how many pools have RBD mirroring enabled
- how many RGW daemons, zones, and zonegroups are present; which RGW frontends are in use
- aggregate stats about the CRUSH map, like which algorithms are used, how big buckets are, how many rules are defined, and what tunables are in use

If you had telemetry enabled, you will need to re-opt-in with:

```
1. ceph telemetry on
```

You can view exactly what information will be reported first with:

```
1. $ ceph telemetry show      # see everything
2. $ ceph telemetry show basic # basic cluster info (including all of the new info)
```

- Following invalid settings now are not tolerated anymore for the command `ceph osd erasure-code-profile set xxx`.
  - \* invalid m for “reed\_sol\_r6\_op” erasure technique
  - \* invalid m and invalid w for “liber8tion” erasure technique
- New OSD daemon command `dump_recovery_reservations` which reveals the recovery locks held (`in_progress`) and waiting in priority queues.
- New OSD daemon command `dump_scrub_reservations` which reveals the scrub reservations that are held for local (primary) and remote (replica) PGs.
- Previously, `ceph tell mgr ...` could be used to call commands implemented by mgr modules. This is no longer supported. Since luminous, using `tell` has not been necessary: those same commands are also accessible without the `tell mgr` portion (e.g., `ceph tell mgr influx foo` is the same as `ceph influx foo`. `ceph tell mgr ...` will

now call admin commands—the same set of commands accessible via `ceph daemon ...` when you are logged into the appropriate host.

- The `ceph tell` and `ceph daemon` commands have been unified, such that all such commands are accessible via either interface. Note that ceph-mgr tell commands are accessible via either `ceph tell mgr ...` or `ceph tell mgr.<id> ...`, and it is only possible to send tell commands to the active daemon (the standbys do not accept incoming connections over the network).
- Ceph will now issue a health warning if a RADOS pool has a `pg_num` value that is not a power of two. This can be fixed by adjusting the pool to a nearby power of two:

```
1. ceph osd pool set <pool-name> pg_num <new-pg-num>
```

Alternatively, the warning can be silenced with:

```
1. ceph config set global mon_warn_on_pool_pg_num_not_power_of_two false
```

- The format of MDSS in `ceph fs dump` has changed.
- The `mds_cache_size` config option is completely removed. Since luminous, the `mds_cache_memory_limit` config option has been preferred to configure the MDS's cache limits.
- The `pg_autoscale_mode` is now set to `on` by default for newly created pools, which means that Ceph will automatically manage the number of PGs. To change this behavior, or to learn more about PG autoscaling, see [Autoscaling placement groups](#). Note that existing pools in upgraded clusters will still be set to `warn` by default.
- The pool parameter `target_size_ratio`, used by the pg autoscaler, has changed meaning. It is now normalized across pools, rather than specifying an absolute ratio. For details, see [Autoscaling placement groups](#). If you have set target size ratios on any pools, you may want to set these pools to autoscale `warn` mode to avoid data movement during the upgrade:

```
1. ceph osd pool set <pool-name> pg_autoscale_mode warn
```

- The `upmap_max_iterations` config option of mgr/balancer has been renamed to `upmap_max_optimizations` to better match its behaviour.
- `mClockClientQueue` and `mClockClassQueue` OpQueue implementations have been removed in favor of a single `mClockScheduler` implementation of a simpler OSD interface. Accordingly, the `osd_op_queue_mclock*` family of config options has been removed in favor of the `osd_mclock_scheduler*` family of options.
- The config subsystem now searches dot ('.') delimited prefixes for options. That

means for an entity like `client.foo.bar`, its overall configuration will be a combination of the global options, `client`, `client.foo`, and `client.foo.bar`. Previously, only global, `client`, and `client.foo.bar` options would apply. This change may affect the configuration for clients that include a `.` in their name.

- MDS default cache memory limit is now 4GB.
- The behaviour of the `-o` argument to the rados tool has been reverted to its original behaviour of indicating an output file. This reverts it to a more consistent behaviour when compared to other tools. Specifying object size is now accomplished by using an upper-case O `-O`.
- In certain rare cases, OSDs would self-classify themselves as type ‘nvme’ instead of ‘hdd’ or ‘ssd’. This appears to be limited to cases where BlueStore was deployed with older versions of ceph-disk, or manually without ceph-volume and LVM. Going forward, the OSD will limit itself to only ‘hdd’ and ‘ssd’ (or whatever device class the user manually specifies).
- RGW: a mismatch between the bucket notification documentation and the actual message format was fixed. This means that any endpoints receiving bucket notification, will now receive the same notifications inside an JSON array named ‘Records’. Note that this does not affect pulling bucket notification from a subscription in a ‘pubsub’ zone, as these are already wrapped inside that array.
- The configuration value `osd_calc_pg_upmaps_max_stddev` used for upmap balancing has been removed. Instead use the mgr balancer config `upmap_max_deviation` which now is an integer number of PGs of deviation from the target PGs per OSD. This can be set with a command like `ceph config set mgr/balancer/upmap_max_deviation 2`. The default `upmap_max_deviation` is 1. There are situations where crush rules would not allow a pool to ever have completely balanced PGs. For example, if crush requires 1 replica on each of 3 racks, but there are fewer OSDs in one of the racks. In those cases, the configuration value can be increased.
- MDS daemons can now be assigned to manage a particular file system via the new `mds_join_fs` option. The monitors will try to use only MDS for a file system with `mds_join_fs` equal to the file system name (strong affinity). Monitors may also deliberately failover an active MDS to a standby when the cluster is otherwise healthy if the standby has stronger affinity.
- RGW Multisite: A new fine grained bucket-granularity policy configuration system has been introduced and it supersedes the previous coarse zone sync configuration (specifically the `sync_from` and `sync_from_all` fields in the zonegroup configuration. New configuration should only be configured after all relevant zones in the zonegroup have been upgraded.
- RGW S3: Support has been added for BlockPublicAccess set of APIs at a bucket level, currently blocking/ignoring public acls & policies are supported. User/Account level APIs are planned to be added in the future

- RGW: The default number of bucket index shards for new buckets was raised from 1 to 11 to increase the amount of write throughput for small buckets and delay the onset of dynamic resharding. This change only affects new deployments/zones. To change this default value on existing deployments, use `radosgw-admin zonegroup modify --bucket-index-max-shards=11`. If the zonegroup is part of a realm, the change must be committed with `radosgw-admin period update --commit` - otherwise the change will take effect after radosgws are restarted.

# Changelog

- .gitignore: add more stuff ([pr#29568](#), Volker Theile)
- async/dpdk: fix compile errors from ceph::mutex update ([pr#30066](#), yehu)
- bluestore,build/ops,common,rgw: Enable \_GLIBCXX\_ASSERTIONS and fix unittest problems ([pr#32387](#), Samuel Just)
- bluestore,cephfs,common,core,mgr,mon,rbd,rgw: src/: s/Mutex/ceph::mutex/ ([pr#29113](#), Kefu Chai)
- bluestore,common,core,mgr,rbd: common/RefCountedObj: cleanup con/des ([pr#29672](#), Patrick Donnelly)
- bluestore,common,core,rgw: common, \\*: kill the bl::last\_p member. Use iterator instead ([pr#32831](#), Radoslaw Zarzynski)
- bluestore,common: os/bluestore: s/align\_down/p2align/ ([pr#29379](#), Kefu Chai)
- bluestore,core: common/options: Set bluestore min\_alloc size to 4K ([pr#30698](#), Mark Nelson)
- bluestore,core: common/options: Set concurrent bluestore rocksdb compactions to 2 ([pr#29027](#), Mark Nelson)
- bluestore,core: mon,osd: only use new per-pool usage stats once \\*all\\* osds are reporting ([pr#28978](#), Sage Weil)
- bluestore,core: os/bluestore,mon: segregate omap keys by pool; report via df ([pr#29292](#), Sage Weil)
- bluestore,core: os/bluestore/BlueFS: explicit check for too-granular allocations ([pr#33027](#), Sage Weil)
- bluestore,core: os/bluestore/bluefs\_types: consolidate contiguous extents ([pr#28821](#), Sage Weil)
- bluestore,core: os/bluestore/KernelDevice: fix RW\_IO\_MAX constant ([pr#29577](#), Sage Weil)
- bluestore,core: os/bluestore: do not set osd\_memory\_target default from cgroup limit ([pr#29581](#), Sage Weil)
- bluestore,core: os/bluestore: drop (semi-broken) nvme automatic class ([pr#31796](#), Sage Weil)
- bluestore,core: os/bluestore: expand lttng tracepoints, improve fio\_ceph\_objectstore backend ([pr#29674](#), Samuel Just)

- bluestore,core: os/bluestore: Keep separate onode cache pinned list ([pr#30964](#), Mark Nelson)
- bluestore,core: os/bluestore: prefix omap of temp objects by real pool ([pr#29717](#), xie xingguo)
- bluestore,core: os/bluestore: Unify on preadv for io\_uring and future refactor ([pr#28025](#), Mark Nelson)
- bluestore,core: os/bluestore: v.2 framework for more intelligent DB space usage ([pr#29687](#), Igor Fedotov)
- bluestore,mgr,rgw: rgw,bluestore: fixes to address failures from check-generated.sh ([pr#29862](#), Kefu Chai)
- bluestore,mon: os/bluestore: create the tail when first set FLAG\_OMAP ([pr#27627](#), Tao Ning)
- bluestore,tools: os/bluestore/bluestore-tool: minor fixes around migrate ([pr#28651](#), Igor Fedotov)
- bluestore,tools: tools/ceph-objectstore-tool: implement onode metadata dump ([pr#27869](#), Igor Fedotov)
- bluestore,tools: tools/ceph-objectstore-tool: introduce list-slow-omap command ([pr#27985](#), Igor Fedotov)
- bluestore: BlueFS: prevent BlueFS::dirty\_files from being leaked when syncing metadata ([pr#30631](#), Xuehan Xu)
- bluestore: bluestore/allocator: Ageing test for bluestore allocators ([pr#22574](#), Adam Kupczyk)
- bluestore: bluestore/bdev: initialize size when creating object ([pr#29968](#), Willem Jan Withagen)
- bluestore: bluestore/bluefs: make accounting resiliant to unlock() ([pr#32584](#), Adam Kupczyk)
- bluestore: common/options.cc: change default value of bluestore\_fsck\_on\_mount\_deep to false ([pr#29408](#), Neha Ojha)
- bluestore: common/options: bluestore 64k min\_alloc\_size for HDD ([pr#32809](#), Sage Weil)
- bluestore: NVMEDevice: Remove the unnecessary aio\_wait in sync read ([pr#33597](#), Ziye Yang)
- bluestore: NVMEDevice: Split the read I/O if the io size is large ([pr#32647](#), Ziye Yang)

- bluestore: os/bluestore/Blue(FS|Store): uint64\_t alloc\_size ([pr#32484](#), Bernd Zeimetz)
- bluestore: os/bluestore/BlueFS: clear newly allocated space for WAL logs ([pr#30549](#), Adam Kupczyk)
- bluestore: os/bluestore/BlueFS: fixed printing stats ([pr#33235](#), Adam Kupczyk)
- bluestore: os/bluestore/BlueFS: less verbose about alloc adjustments ([pr#33512](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: Move bluefs alloc size initialization log message to log level 1 ([pr#29822](#), Vikhyat Umrao)
- bluestore: os/bluestore/BlueFS: replace flush\_log with sync\_metadata ([pr#32563](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: use 64K alloc\_size on the shared device ([pr#29537](#), Sage Weil, Neha Ojha)
- bluestore: os/bluestore/BlueStore.cc: set priorities for compression stats ([pr#31959](#), Neha Ojha)
- bluestore: os/bluestore/spdk: Fix the overflow error of parsing spdk coremask ([pr#32440](#), Hu Ye, Chunsong Feng, luo rixin)
- bluestore: os/bluestore: Actually wait until completion in write\_sync ([pr#26909](#), Vitaliy Filippov)
- bluestore: os/bluestore: add bluestore\_bluefs\_max\_free; smooth space balancing a bit ([pr#30231](#), xie xingguo)
- bluestore: os/bluestore: add slow op detection for collection\_listing ([issue#40741](#), [pr#29085](#), Igor Fedotov)
- bluestore: os/bluestore: allocate Task on stack ([pr#33358](#), Jun Su)
- bluestore: os/bluestore: apply garbage collection against excessive blob count growth ([pr#28229](#), Igor Fedotov)
- bluestore: os/bluestore: AVL-tree & extent - based space allocator ([pr#30897](#), Adam Kupczyk, xie xingguo, Kefu Chai)
- bluestore: os/bluestore: avoid length overflow in extents returned by Stupid ([issue#40703](#), [pr#28945](#), Igor Fedotov)
- bluestore: os/bluestore: avoid race between split\_cache and get/put pin/unpin ([pr#32665](#), Sage Weil)
- bluestore: os/bluestore: avoid unnecessary notify ([pr#29345](#), Jianpeng Ma)
- bluestore: os/bluestore: be more verbose doing bluefs log replay ([pr#27615](#), Igor

Fedotov)

- bluestore: os/bluestore: bluefs\_preextend\_wal\_files=true ([pr#28322](#), Sage Weil)
- bluestore: os/bluestore: call fault\_range prior to looking for blob to reuse ([pr#27444](#), Igor Fedotov)
- bluestore: os/bluestore: check bluefs allocations on log replay ([pr#31513](#), Igor Fedotov)
- bluestore: os/bluestore: check return value of func \_open\_db\_and\_around ([pr#27477](#), Jianpeng Ma)
- bluestore: os/bluestore: cleanup around allocator calls ([pr#29068](#), Igor Fedotov)
- bluestore: os/bluestore: cleanups ([pr#30737](#), Kefu Chai)
- bluestore: os/bluestore: consolidate extents from the same device only ([pr#31621](#), Igor Fedotov)
- bluestore: os/bluestore: correctly measure deferred writes into new blobs ([issue#38816](#), [pr#27789](#), Sage Weil)
- bluestore: os/bluestore: deferred IO notify and locking optimization ([pr#29522](#), Jianpeng Ma)
- bluestore: os/bluestore: do not check osd\_max\_object\_size in \_open\_path() ([pr#26176](#), Igor Fedotov)
- bluestore: os/bluestore: do not mark per\_pool\_omap updated unless we fixed it ([pr#31167](#), Sage Weil)
- bluestore: os/bluestore: dont round\_up\_to in apply\_for\_bitset\_range ([pr#31903](#), Jianpeng Ma)
- bluestore: os/bluestore: dump onode before no available blob id abort ([pr#27911](#), Igor Fedotov)
- bluestore: os/bluestore: dump onode that has too many spanning blobs ([pr#28010](#), Igor Fedotov)
- bluestore: os/bluestore: fix >2GB writes ([pr#27871](#), Sage Weil, kungf)
- bluestore: os/bluestore: fix bitmap allocator issues ([pr#26939](#), Igor Fedotov)
- bluestore: os/bluestore: fix duplicate allocations in bmap allocator ([issue#40080](#), [pr#28496](#), Igor Fedotov)
- bluestore: os/bluestore: fix duplicative and misleading debug in KernelDevice::open() ([pr#28630](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: fix for FreeBSD iocb structure ([pr#27458](#), Willem Jan

Withagen)

- bluestore: os/bluestore: fix invalid stray shared blob detection in fsck ([pr#30616](#), Igor Fedotov)
- bluestore: os/bluestore: fix missing discard in BlueStore::\_kv\_sync\_thread ([pr#27843](#), Junhui Tang)
- bluestore: os/bluestore: fix origin reference in logging slow ops ([pr#27951](#), Igor Fedotov)
- bluestore: os/bluestore: fix out-of-bound access in bmap allocator ([pr#27691](#), Igor Fedotov)
- bluestore: os/bluestore: fix per-pool omap repair ([pr#32925](#), Igor Fedotov)
- bluestore: os/bluestore: fix space balancing overflow ([pr#30255](#), xie xingguo)
- bluestore: os/bluestore: fix wakeup bug ([pr#31931](#), Jianpeng Ma)
- bluestore: os/bluestore: introduce legacy statfs and dev size mismatch alerts ([pr#27519](#), Sage Weil, Igor Fedotov)
- bluestore: os/bluestore: introduce new io\_uring IO engine ([pr#27392](#), Roman Penyaev)
- bluestore: os/bluestore: its better to erase spanning blob once ([pr#29238](#), Xiangyang Yu)
- bluestore: os/bluestore: load OSD all compression settings unconditionally ([issue#40480](#), [pr#28688](#), Igor Fedotov)
- bluestore: os/bluestore: log allocation stats on a daily basis ([pr#33565](#), Igor Fedotov)
- bluestore: os/bluestore: memorize layout of BlueFS on management ([pr#30593](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: Merge deferred\_finisher and finisher ([pr#29623](#), Jianpeng Ma)
- bluestore: os/bluestore: minor improvements/cleanup around allocator ([pr#29738](#), Igor Fedotov)
- bluestore: os/bluestore: more aggressive deferred submit when onode trim skipping ([issue#21531](#), [pr#25697](#), Zengran Zhang)
- bluestore: os/bluestore: more smart allocator dump when lacking space for bluefs ([issue#40623](#), [pr#28845](#), Igor Fedotov)
- bluestore: os/bluestore: new bluestore\_debug\_enforce\_settings option ([pr#27132](#), Igor Fedotov)

- bluestore: os/bluestore: no need protected by OpSequencer::qlock ([pr#29488](#), Jianpeng Ma)
- bluestore: os/bluestore: no need to add tail length (revert PR#29185) ([pr#29465](#), Xiangyang Yu)
- bluestore: os/bluestore: print correctly info ([pr#29939](#), Jianpeng Ma)
- bluestore: os/bluestore: print error if spdk\_nvme\_ns\_cmd\_writev() fails ([pr#31932](#), NancySu05)
- bluestore: os/bluestore: proper locking for BlueFS prefetching ([pr#29012](#), Igor Fedotov)
- bluestore: os/bluestore: reduce wakeups ([pr#29130](#), Jianpeng Ma)
- bluestore: os/bluestore: Refactor Bluestore Caches ([pr#28597](#), Mark Nelson)
- bluestore: os/bluestore: remove unused arg to \_get\_deferred\_op() ([issue#40918](#), [pr#29320](#), Sage Weil)
- bluestore: os/bluestore: remove unused \_tune\_cache\_size() method declaration ([pr#29393](#), Igor Fedotov)
- bluestore: os/bluestore: restore and fix bug with onode cache pinning ([pr#31778](#), Josh Durgin)
- bluestore: os/bluestore: revert cache pinned list ([pr#31180](#), Sage Weil)
- bluestore: os/bluestore: set STATE\_KV\_SUBMITTED properly ([pr#30753](#), Igor Fedotov)
- bluestore: os/bluestore: show device name in osd metadata output ([pr#28107](#), Igor Fedotov)
- bluestore: os/bluestore: silence StupidAllocator reorder warning ([pr#29866](#), Jos Collin)
- bluestore: os/bluestore: simplify multithreaded shallow fsck ([pr#31473](#), Igor Fedotov)
- bluestore: os/bluestore: simplify per-pool-stat config options ([pr#30350](#), Sage Weil, Igor Fedotov)
- bluestore: os/bluestore: support RocksDB prefetch in buffered read mode ([issue#36482](#), [pr#27782](#), Igor Fedotov)
- bluestore: os/bluestore: tiny tracepoints improvement ([pr#31669](#), Adam Kupczyk)
- bluestore: os/bluestore: upgrade legacy omap to per-pool format automatically ([pr#32758](#), Igor Fedotov)
- bluestore: os/bluestore: verify disk layout of BlueFS ([issue#25098](#), [pr#30109](#),

Radoslaw Zarzynski)

- bluestore: os/bluestore:fix two calculation bugs ([pr#29185](#), Xiangyang Yu)
- bluestore: os/ceph-bluestore-tool: bluefs-bdev-expand asserts if no WAL ([pr#27445](#), Igor Fedotov)
- bluestore: os/objectstore: add new op OP\_CREATE for create a new object ([pr#26251](#), Jianpeng Ma)
- bluestore: Revert os/bluestore: add kv\_drain\_preceding\_waiters indicate drain\_preceding. ([pr#31503](#), Sage Weil)
- bluestore: test/fio: handle nullptr when parsing throttle params ([pr#31681](#), Igor Fedotov)
- bluestore: [bluestore][tools] Inspect allocations in bluestore ([pr#29425](#), Adam Kupczyk)
- build(deps): bump lodash from 4.17.11 to 4.17.13 in /src/pybind/mgr/dashboard/frontend ([pr#29192](#), dependabot[bot])
- build/ops,cephfs,common,core,rbd: Fix big-endian handling ([pr#30079](#), Ulrich Weigand)
- build/ops,cephfs: mgr/ssh: make mds add work ([pr#31059](#), Sage Weil)
- build/ops,common,core: common, include: bump the version of ceph::buffers C++ API ([pr#33373](#), Radoslaw Zarzynski)
- build/ops,common,mgr: python-common: Python common package ([pr#28915](#), Kefu Chai, Sebastian Wagner)
- build/ops,common,rgw: rgw, common, build: drop NSS support ([pr#27834](#), Radoslaw Zarzynski)
- build/ops,core,rbd: Windows support [part 1] ([pr#31981](#), Lucian Petrut, Alin Gabriel Serdean)
- build/ops,core: ceph-crash: use client.crash[.host] to post, and provision keys via mgr/ssh + ceph-daemon ([pr#30734](#), Sage Weil)
- build/ops,core: debian: fix ceph-mgr-modules-core files ([pr#33468](#), Sage Weil)
- build/ops,core: os/bluestore: fix pmem osd build problem ([pr#28761](#), Peterson, Scott, Li, Xiaoyan)
- build/ops,core: qa: stop testing on 16.04 xenial ([pr#28943](#), Sage Weil)
- build/ops,mgr: mgr/diskprediction\_local: Replaced old models and updated predictor ([pr#29437](#), Karanraj Chauhan)

- build/ops,mgr: systemd: ceph-mgr: set MemoryDenyWriteExecute to false ([issue#39628](#), [pr#28023](#), Ricardo Dias)
- build/ops,pybind: cmake, pybind: fix build on armhf ([pr#28843](#), Kefu Chai)
- build/ops,rbd: rpm,deb: fix python dateutil module dependency ([pr#33624](#), Mykola Golub)
- build/ops,rgw: build/rgw: unittest\_rgw\_dmclock\_scheduler does not need Boost\_LIBRARIES ([pr#27466](#), Willem Jan Withagen)
- build/ops,rgw: install-deps.sh, cmake: use boost 1.72 on bionic ([pr#32391](#), Kefu Chai)
- build/ops,tests: ceph-daemon: a few fixes; functional test ([pr#31094](#), Sage Weil)
- build/ops,tests: googletest: pick up change to suppress CMP0048 warning ([pr#29471](#), Kefu Chai)
- build/ops,tests: install-deps.sh,deb,rpm: move python-saml deps into debian/control anxe2x80xa6 ([pr#29840](#), Kefu Chai)
- build/ops,tools: src/script/credits.sh - switch to bash ([pr#32736](#), Kai Wagner)
- build/ops,tools: vstart: Now all OSDs are starting in parallel. Use -no-parallel to revert to sequential ([pr#31732](#), Adam Kupczyk)
- build/ops: .github/stale.yml: warn at 60, close at 90; adjust message ([pr#24744](#), Lenz Grimmer, Sage Weil)
- build/ops: admin/build-doc: keep-going when finding warnings ([pr#27050](#), Abhishek Lekshmanan)
- build/ops: build-doc: allow building docs on fedora 30 ([pr#30136](#), Yuval Lifshitz)
- build/ops: build-integration-branch: s/prefix/postfix/ ([pr#32303](#), Kefu Chai)
- build/ops: build: add static analysis targets ([pr#31579](#), Yuval Lifshitz)
- build/ops: build: FreeBSD does not have /etc/os-release ([pr#26731](#), Willem Jan Withagen)
- build/ops: ceph-daemon: a couple fixes ([pr#31060](#), Sage Weil)
- build/ops: ceph-daemon: add a logrotate.d file for each cluster ([pr#30882](#), Sage Weil)
- build/ops: ceph-daemon: deploy ceph daemons with podman and systemd ([pr#30603](#), Sage Weil)
- build/ops: ceph-daemon: fix logrotate su line ([pr#31823](#), Sage Weil)

- build/ops: ceph-daemon: misc improvements ([pr#30826](#), Sage Weil)
- build/ops: ceph-daemon: use /usr/bin/python, not /usr/bin/env python ([pr#31318](#), Sage Weil)
- build/ops: ceph.spec.in: add missing python-yaml dependency for mgr-k8sevents ([pr#31178](#), Kefu Chai)
- build/ops: ceph.spec.in: add runtime deps for mgr-diskprediction-cloud ([pr#32232](#), Kefu Chai)
- build/ops: ceph.spec.in: always depends on python3.6-pyOpenSSL ([pr#32317](#), Kefu Chai)
- build/ops: ceph.spec.in: Drop systemd BuildRequires in case of building for SUSE ([pr#28884](#), Dominique Leuenberger)
- build/ops: ceph.spec.in: enable amqp\_endpoint on RHEL8 by default ([pr#31143](#), Brad Hubbard)
- build/ops: ceph.spec.in: fix Cython package dependency for Fedora ([pr#30590](#), Jeff Layton)
- build/ops: ceph.spec.in: fix make check deps for centos8 ([pr#32798](#), Alfonso Martxc3xadnez)
- build/ops: ceph.spec.in: fix python coverage dependency for non-rhel distros ([pr#33361](#), Kiefer Chang)
- build/ops: ceph.spec.in: fix python3 dependencies in centos7 ([pr#32775](#), liushi)
- build/ops: ceph.spec.in: grafana-dashboards package depends on grafana ([pr#28228](#), Jan Fajerski)
- build/ops: ceph.spec.in: move distro-conditional deps to dedicated section ([pr#32080](#), Nathan Cutler)
- build/ops: ceph.spec.in: package prometheus default alerts for SUSE ([pr#27996](#), Jan Fajerski)
- build/ops: ceph.spec.in: pin to gcc-c++-8.2.1 ([pr#28859](#), Kefu Chai)
- build/ops: ceph.spec.in: re-enable make check deps for el8 ([pr#32412](#), Kefu Chai)
- build/ops: ceph.spec.in: reserve more memory per build job ([pr#30126](#), Dan van der Ster)
- build/ops: ceph.spec.in: s/pkgversion/version\_nodots/ ([pr#30036](#), Kefu Chai)
- build/ops: ceph.spec.in: use g++ >= 8.3.1-3.1 ([pr#30088](#), Kefu Chai)
- build/ops: ceph.spec.in: Use pkgconfig() style BuildRequires for udev/libudev-

- devel ([pr#32933](#), Dominique Leuenberger)
- build/ops: ceph.spec.in: use python3 to bytecompile .py files ([pr#32608](#), Kefu Chai)
- build/ops: ceph.spec: Recommend (but do not require) podman ([pr#33221](#), Sage Weil)
- build/ops: ceph\_release: octopus rc 15.1.0 ([pr#32623](#), Sage Weil)
- build/ops: cmake,crimson: pick up latest seastar ([pr#27088](#), Kefu Chai)
- build/ops: cmake,run-make-check.sh: disable SPDK by default ([pr#29728](#), Kefu Chai)
- build/ops: cmake/Boost: Fix python3 version ([pr#32344](#), Kotresh HR)
- build/ops: cmake/FindRocksDB: fix IMPORTED\_LOCATION for ROCKSDB\_LIBRARIES ([pr#26813](#), dudengke)
- build/ops: cmake/modules/GetGitRevisionDescription: update to work with git-worktree ([pr#30772](#), Sage Weil)
- build/ops: cmake/modules: replace ; with in compile flags ([pr#28339](#), Kefu Chai)
- build/ops: CMakeLists: add std::move warnings in gcc9 ([pr#27569](#), Patrick Donnelly)
- build/ops: crimson: clang related cleanups ([pr#33680](#), Kefu Chai)
- build/ops: crimson: fix build seastar with dpdk ([pr#31426](#), Yingxin Cheng)
- build/ops: deb,rpm,doc: s/plugin/module/ ([pr#33435](#), Kefu Chai)
- build/ops: debian/: use ceph-osd for packaging crimson-osd ([pr#28535](#), Kefu Chai)
- build/ops: debian/control: add python-routes dependency for dashboard ([pr#28835](#), Paul Emmerich)
- build/ops: debian/control: Build-Depends on g++ ([pr#30410](#), Kefu Chai)
- build/ops: debian/control: fix Build-Depends ([pr#29913](#), Kefu Chai)
- build/ops: debian/radosgw.install: correct path to libradosgw.so\\* ([pr#32539](#), Kefu Chai)
- build/ops: debian/rules: run dh\_python2 with ceph-daemon ([pr#31313](#), Kefu Chai)
- build/ops: debian: modules-core replaces and breaks older ceph-mgr ([pr#33501](#), Kefu Chai)
- build/ops: debian: remove dup ceph-fuse line ([pr#28788](#), huangjun)
- build/ops: dmclock: pick up change to use specified C++ settings if any ([pr#30113](#), Kefu Chai)

- build/ops: do\_cmake.sh: Add a heading to the minimal config ([pr#28776](#), Brad Hubbard)
- build/ops: do\_cmake.sh: Add CEPH\_GIT\_DIR ([pr#30863](#), Matthew Oliver)
- build/ops: do\_cmake.sh: bail out if something goes wrong ([pr#33016](#), Kefu Chai)
- build/ops: do\_cmake.sh: enable amqp and rdma for EL8 ([pr#30974](#), Kefu Chai)
- build/ops: do\_cmake.sh: optionally specify build dir with \$BUILD\_DIR env var ([pr#29786](#), Yuval Lifshitz)
- build/ops: do\_cmake.sh: remove -DCMAKE\_BUILD\_TYPE=Debug from cmake options ([pr#30250](#), Kefu Chai)
- build/ops: do\_cmake.sh: use bash ([issue#39981](#), [pr#28181](#), Nathan Cutler)
- build/ops: do\_cmake: Warn user about slow debug performance only for not set ([pr#31113](#), Junyoung Sung)
- build/ops: do\_freebsd.sh: update build scripts to resemble Jenkins scripts ([pr#29400](#), Willem Jan Withagen)
- build/ops: dpdk: drop dpdk submodule ([issue#24032](#), [pr#33001](#), Kefu Chai)
- build/ops: fix build fail related to PYTHON\_EXECUTABLE variable ([pr#30199](#), Ilsoo Byun)
- build/ops: github: display phrase for signed-off check ([pr#29890](#), Ernesto Puerta)
- build/ops: install-dep,rpm: use devtools-8 on amd64 ([issue#38892](#), [pr#27134](#), Kefu Chai)
- build/ops: install-deps, rpm: use python\_provide macro and cleanups ([pr#30830](#), Kefu Chai)
- build/ops: install-deps,rpm,do\_cmake: build on RHEL/CentOS 8 ([pr#30630](#), Kefu Chai)
- build/ops: install-deps.sh,src: drop python2 support ([pr#31525](#), Kefu Chai)
- build/ops: install-deps.sh: Actually set gpgcheck to false ([pr#33591](#), Brad Hubbard)
- build/ops: install-deps.sh: add EPEL repo for non-x86\_64 archs as well ([pr#30557](#), Kefu Chai, Nathan Cutler)
- build/ops: install-deps.sh: add kens copr repo for el8 build ([pr#32324](#), Kefu Chai)
- build/ops: install-deps.sh: add option to skip prebuilt boost-\* pkgs installation ([pr#27776](#), Jun He)

- build/ops: install-deps.sh: add support for Ubuntu Disco Dingo ([pr#30405](#), Patrick Seidensal)
- build/ops: install-deps.sh: download wheel using pip wheel ([pr#29903](#), Kefu Chai)
- build/ops: install-deps.sh: enable PowerTool repo for EL8 ([pr#30656](#), Kefu Chai)
- build/ops: install-deps.sh: fix typo for krb5 on FreeBSD ([pr#28269](#), Thomas Johnson)
- build/ops: install-deps.sh: install binutils 2.28 for xenial ([pr#31601](#), Kefu Chai)
- build/ops: install-deps.sh: install libboost-test for seastar ([pr#28015](#), Kefu Chai)
- build/ops: install-deps.sh: install python2-{virtualenv,devel} on SUSE if needed ([pr#32153](#), Nathan Cutler)
- build/ops: install-deps.sh: install \\*rpm-macros ([issue#39164](#), [pr#27524](#), Kefu Chai)
- build/ops: install-deps.sh: install python\\*-devel for python\\*rpm-macros ([pr#30190](#), Kefu Chai)
- build/ops: install-deps.sh: only prepare wheels for make check ([pr#29912](#), Kefu Chai)
- build/ops: install-deps.sh: use chacra for cmake repo ([pr#29475](#), Kefu Chai)
- build/ops: install-deps.sh: Use dnf for rhel/centos 8 ([pr#31144](#), Brad Hubbard)
- build/ops: install-deps.sh: use gcc-8 on xenial and trusty ([pr#28094](#), Kefu Chai)
- build/ops: install-deps.sh: use GCC-9 on bionic ([pr#28454](#), Kefu Chai)
- build/ops: install-deps.sh: use sepia/lab-extra/8 ([pr#31238](#), Kefu Chai)
- build/ops: install-deps: do not install if rpm already installed ([pr#30612](#), Kefu Chai)
- build/ops: install-deps: enable homebrew repos for RHEL8 ([pr#33905](#), Kefu Chai, Dan Mick)
- build/ops: install-deps: revert 47d4351d ([pr#30122](#), Kefu Chai)
- build/ops: make patch build dependency explicit ([issue#40175](#), [pr#28414](#), Nathan Cutler)
- build/ops: make perf\_async\_msgr link jemalloc/tcmalloc ([pr#28039](#), Jianpeng Ma)
- build/ops: make-dist: Bump Node.js to v10.18.1 ([pr#33059](#), Tiago Melo)

- build/ops: make-dist: default to no dashboard frontend build parallelism ([pr#32037](#), Nathan Cutler)
- build/ops: make-dist: drop Python 2/3 autoselect ([pr#27792](#), Nathan Cutler)
- build/ops: make-dist: set version number only once ([pr#26281](#), Nathan Cutler)
- build/ops: mgr/dashboard: Prevent angular of getting stuck during installation ([pr#29929](#), Tiago Melo)
- build/ops: mgr/rook: Make use of rook-client-python when talking to Rook ([pr#29427](#), Sebastian Wagner)
- build/ops: pybind/mgr/CMakeLists: exclude tox.ini, requirements.txt from install ([pr#31577](#), Sage Weil)
- build/ops: pybind/mgr: Exclude tests/ ([pr#31671](#), Sebastian Wagner)
- build/ops: pybind/mgr: Rename orchestrator\_cli to orchestrator ([pr#32817](#), Sebastian Wagner)
- build/ops: qa/tasks/ceph\_deploy: do not rely on ceph-create-keys ([pr#29002](#), Sage Weil)
- build/ops: Revert dpdk: drop dpdk submodule ([pr#32992](#), David Galloway)
- build/ops: rpm,cmake: use specified python3 version if any ([pr#27358](#), Kefu Chai)
- build/ops: rpm,deb: package always-enabled plugins in a separated package ([pr#33422](#), Kefu Chai)
- build/ops: rpm,deb: python-requests is not needed for ceph-common ([pr#30420](#), luo.runbing)
- build/ops: rpm,debian,install-deps: package crimson-osd ([pr#28428](#), Kefu Chai)
- build/ops: rpm,etc/sysconfig: remove SuSEfirewall2 support ([issue#40738](#), [pr#28957](#), Matthias Gerstner)
- build/ops: rpm/cephadm: move HOMEDIR to /var/lib and make scriptlets idempotent on SUSE ([pr#32212](#), Nathan Cutler)
- build/ops: rpm: add cmake\_verbose\_logging switch ([pr#32805](#), Nathan Cutler)
- build/ops: rpm: add Provides: python3-\* for python packages and cleanup ([pr#27468](#), Kefu Chai)
- build/ops: rpm: add rpm-build to SUSE-specific make check deps ([pr#32083](#), Nathan Cutler)
- build/ops: rpm: always build ceph-test package ([pr#29685](#), Nathan Cutler)

- build/ops: rpm: define weak\_deps for el8 ([pr#33229](#), Kefu Chai)
- build/ops: rpm: Disable LTO in spec when being used ([issue#39974](#), [pr#28170](#), Martin Lixc5xa1ka)
- build/ops: rpm: drop vim-specific header ([pr#32331](#), Nathan Cutler)
- build/ops: rpm: enable devtoolset-8 on aarch64 also ([issue#38892](#), [pr#27333](#), Kefu Chai)
- build/ops: rpm: fdupes in SUSE builds to conform with packaging guidelines ([issue#40973](#), [pr#29346](#), Nathan Cutler)
- build/ops: rpm: fix rhel <= 7 conditional ([pr#27045](#), Nathan Cutler)
- build/ops: rpm: fix up a specfile syntax error ([pr#33066](#), Greg Farnum)
- build/ops: rpm: have pybind RPMs provide/obsolete their python2 predecessors ([issue#40099](#), [pr#28352](#), Nathan Cutler)
- build/ops: rpm: immutable-object-cache related changes ([pr#27150](#), Kefu Chai)
- build/ops: rpm: improve ceph-mgr plugin package summaries ([issue#40974](#), [pr#29347](#), Nathan Cutler)
- build/ops: rpm: make librados2, libcephfs2 own (create) /etc/ceph ([pr#30975](#), Nathan Cutler)
- build/ops: rpm: put librgw lttng SOS in the librgw-devel package ([issue#40975](#), [pr#29349](#), Nathan Cutler)
- build/ops: rpm: refrain from building ceph-resource-agents on SLE ([pr#27046](#), Nathan Cutler)
- build/ops: rpm: Relax the selinux policy version for centos builds ([pr#32700](#), Boris Ranto)
- build/ops: rpm: s/devtoolset-7/devtoolset-8/ ([pr#27183](#), Kefu Chai)
- build/ops: rpm: use python 3.6 as the default python3 ([pr#27417](#), Kefu Chai)
- build/ops: rpm: use python3.4 on RHEL7 by default ([pr#27407](#), Kefu Chai)
- build/ops: rpm: use Recommends on fedora also ([pr#26819](#), Kefu Chai)
- build/ops: run npm ci with a one-hour timeout ([pr#28994](#), Nathan Cutler)
- build/ops: run-make-check.sh: extract run-make.sh ([pr#30184](#), Kefu Chai)
- build/ops: run-make-check.sh: run sudo with absolute path ([pr#29753](#), Kefu Chai)
- build/ops: run-make-check.sh: WITH\_SEASTAR on demand ([pr#33723](#), Kefu Chai)

- build/ops: script, doc: add gen-corpus.sh ([pr#28950](#), Kefu Chai)
- build/ops: script/build-integration-branch: Add usage ([pr#32293](#), Sebastian Wagner)
- build/ops: script/run-make.sh: do not pass cmake options twice ([pr#30318](#), Kefu Chai)
- build/ops: script/run\_tox.sh: Dont overwrite the build dir ([pr#29925](#), Sebastian Wagner)
- build/ops: script: remove dep-report.sh ([pr#29296](#), Kefu Chai)
- build/ops: scripts: ceph\_dump\_log.py ([pr#21729](#), Brad Hubbard)
- build/ops: seastar: pickup change to add pthread linkage ([pr#33453](#), Kefu Chai)
- build/ops: spec, debian: cephadm requires lvm2 ([pr#32323](#), Sebastian Wagner)
- build/ops: spec,debian: ceph-mgr-ssh depends on openssh{-client{s}} ([pr#31806](#), Sebastian Wagner)
- build/ops: spec: add missing python3-pyyaml ([pr#33387](#), Sebastian Wagner)
- build/ops: spec: Podman (temporarily) requires apparmor-abstractions on suse ([pr#33850](#), Sebastian Wagner)
- build/ops: src/CMakeLists: remove leading v from git describe version ([pr#31387](#), Sage Weil)
- build/ops: test/fio: bump to fio-3.15 ([pr#31544](#), Igor Fedotov)
- build/ops: test: only compile ceph\_test\_bmap\_alloc\_replay WITH\_BLUESTORE ([pr#31306](#), Willem Jan Withagen)
- build/ops: vstart: Remove duplicate option -N ([pr#31917](#), Kotresh HR)
- ceph-crash: use ceph-crash as logger name ([pr#30989](#), Kefu Chai)
- ceph-daemon -> cephadm, mgr/ssh -> mgr/cephadm ([pr#32193](#), Sage Weil)
- ceph-daemon,mgr/ssh: add check-host ([pr#31795](#), Sage Weil)
- ceph-daemon: -v|-verbose, not -d|-debug ([pr#31583](#), Sage Weil)
- ceph-daemon: a few more py2 compatibility hacks ([pr#31264](#), Sage Weil)
- ceph-daemon: add additional debug logging ([pr#31837](#), Michael Fritch)
- ceph-daemon: Add basic mypy support ([pr#31609](#), Thomas Bechtold)
- ceph-daemon: add explicit pull at bootstrap start ([pr#31478](#), Sage Weil)

- ceph-daemon: Add more type hints ([pr#31631](#), Thomas Bechtold)
- ceph-daemon: add osd create test ([pr#31679](#), Michael Fritch)
- ceph-daemon: add standalone adopt tests ([pr#31486](#), Michael Fritch)
- ceph-daemon: add -base-dir arg to adopt command ([pr#31487](#), Michael Fritch)
- ceph-daemon: add -legacy-dir arg to ls command ([pr#31585](#), Michael Fritch)
- ceph-daemon: Allow env var for setting the used image ([pr#31913](#), Thomas Bechtold)
- ceph-daemon: append newline before public key string ([pr#31788](#), Ricardo Dias)
- ceph-daemon: behave on rm-cluster when legacy dirs exist and ceph isn't installed ([pr#31499](#), Sage Weil)
- ceph-daemon: bootstrap: make -output-\\* args optional ([pr#31695](#), Sage Weil)
- ceph-daemon: ceph/daemon-base:latest-master-devel ([pr#31507](#), Sage Weil)
- ceph-daemon: clean-up tempfiles on EXIT ([pr#32052](#), Michael Fritch)
- ceph-daemon: combine SUDO and ARGS into a single var ([pr#32138](#), Michael Fritch)
- ceph-daemon: configure firewalld for new daemons ([pr#31869](#), Sage Weil)
- ceph-daemon: consolidate NamedTemporaryFile logic ([pr#31908](#), Michael Fritch)
- ceph-daemon: create ~/.ssh if not exist ([pr#31315](#), Kefu Chai)
- ceph-daemon: customize the bash prompt for shell + enter ([pr#31498](#), Sage Weil)
- ceph-daemon: do not pass -it unless it is an interactive shell ([pr#31181](#), Sage Weil)
- ceph-daemon: do not relabel system directories ([pr#31321](#), Sage Weil)
- ceph-daemon: dont deref symlinks during chown ([pr#32137](#), Michael Fritch)
- ceph-daemon: enable dashboard during bootstrap ([pr#31464](#), Sage Weil)
- ceph-daemon: fix bootstrap ownership of tmp monmap file ([pr#32097](#), Sage Weil)
- ceph-daemon: fix extract\_uid\_gid ([pr#31832](#), Sage Weil)
- ceph-daemon: fix firewalld error case ([pr#32096](#), Sage Weil)
- ceph-daemon: Fix handling for symlinks on python2 ([pr#31838](#), Michael Fritch)
- ceph-daemon: fix os.mkdir call ([pr#31320](#), Sage Weil)
- ceph-daemon: fix pod stop ([pr#32157](#), Sage Weil)

- ceph-daemon: fix prompt ([pr#31603](#), Sage Weil)
- ceph-daemon: fix standalone adopt OSD test ([pr#31772](#), Sage Weil, Michael Fritch)
- ceph-daemon: fix traceback during ls command ([pr#31439](#), Michael Fritch)
- ceph-daemon: fix version field for legacy ls ([pr#31443](#), Michael Fritch)
- ceph-daemon: fix systemctl is-enabled bool ([pr#31870](#), Michael Fritch)
- ceph-daemon: infer fsid for some commands ([pr#31702](#), Michael Fritch)
- ceph-daemon: logs command ([pr#31575](#), Sage Weil)
- ceph-daemon: make /var/run/ceph behavior better ([pr#31141](#), Sage Weil)
- ceph-daemon: make infer\_fsid behave when /var/lib/ceph dne ([pr#31831](#), Sage Weil)
- ceph-daemon: make ls log less noisy ([pr#31448](#), Sage Weil)
- ceph-daemon: make mon container privileged ([pr#31476](#), Sage Weil)
- ceph-daemon: make ps1 a raw string ([pr#31540](#), Michael Fritch)
- ceph-daemon: make rm-cluster faster ([pr#31538](#), Sage Weil)
- ceph-daemon: make rm-cluster handle failed unit cleanup ([pr#31365](#), Sage Weil)
- ceph-daemon: Move ceph-daemon executable to own directory ([pr#31467](#), Thomas Bechtold)
- ceph-daemon: nicer errors ([pr#31886](#), Sage Weil, Michael Fritch)
- ceph-daemon: Only run in the \_\_main\_\_ scope ([pr#31458](#), Thomas Bechtold)
- ceph-daemon: only set up /var/run/ceph/\$fsid if it exists ([pr#31341](#), Sage Weil)
- ceph-daemon: only set up crash dir mount if it exists ([pr#31130](#), Sage Weil)
- ceph-daemon: py2 compatibility ([pr#31168](#), Sage Weil)
- ceph-daemon: py2: tolerate whitespace before config key name ([pr#32098](#), Sage Weil)
- ceph-daemon: raise RuntimeError when CephContainer.run() fails ([pr#31328](#), Michael Fritch)
- ceph-daemon: Remove data dir during adopt ([pr#31437](#), Michael Fritch)
- ceph-daemon: remove prepare-host ([pr#32108](#), Sage Weil)
- ceph-daemon: replace podman variables by container ([pr#31618](#), Dimitri Savineau)
- ceph-daemon: seek relative to the start of file ([pr#31892](#), Michael Fritch)

- ceph-daemon: set container\_image during bootstrap ([pr#31445](#), Sage Weil)
- ceph-daemon: set ssh public identity ([pr#31500](#), Sage Weil)
- ceph-daemon: several fsid inference fixes ([pr#31798](#), Sage Weil)
- ceph-daemon: switch default image ([pr#31463](#), Sage Weil)
- ceph-daemon: unmount osd data dir during adopt ([pr#31477](#), Michael Fritch)
- ceph-daemon: use client.admin keyring during bootstrap ([pr#31270](#), Sage Weil)
- ceph-daemon: use -e instead of -env ([pr#31614](#), Michael Fritch)
- ceph-daemon: Use shutil.move to move log files ([pr#31331](#), Michael Fritch)
- ceph-daemon: imp module DeprecationWarning ([pr#32161](#), Michael Fritch)
- ceph-mon: keep v1 address type when explicitly set ([pr#31765](#), Ricardo Dias)
- ceph-object-corpus: forward\_incompat pg\_missing\_item and pg\_missing\_t ([pr#28034](#), lishuhao)
- ceph-volume simple: better detection when type file is not present ([pr#29386](#), Alfredo Deza)
- ceph-volume zap always skips block.db, leaves them around ([issue#40664](#), [pr#28998](#), Alfredo Deza)
- ceph-volume broken assertion errors after pytest changes ([issue#40665](#), [pr#28866](#), Alfredo Deza)
- ceph-volume lvm.zap fix cleanup for db partitions ([issue#40664](#), [pr#28267](#), Dominik Csapak)
- ceph-volume tests add a sleep in tox for slow OSDs after booting ([issue#40619](#), [pr#28836](#), Alfredo Deza)
- ceph-volume tests remove xenial from functional testing ([pr#31159](#), Alfredo Deza)
- ceph-volume tests set the noninteractive flag for Debian ([pr#29804](#), Alfredo Deza)
- ceph-volume-zfs: add the inventory command ([pr#30995](#), Willem Jan Withagen)
- ceph-volume/batch: fail on filtered devices when non-interactive ([pr#31978](#), Jan Fajerski)
- ceph-volume/lvm/activate.py: clarify error message: fsid refers to osd\_fsid ([pr#32351](#), Yaniv Kaul)
- ceph-volume/test: patch VolumeGroups ([pr#31979](#), Jan Fajerski)
- ceph-volume: add Ceph device id to inventory ([pr#31072](#), Sebastian Wagner)

- ceph-volume: add db and wal support to raw mode ([pr#32828](#), Sxc3xa9bastien Han)
- ceph-volume: add methods to pass filters to pvs, vgs and lvs commands ([pr#32242](#), Rishabh Dave)
- ceph-volume: add proper size attribute to partitions ([pr#31492](#), Jan Fajerski)
- ceph-volume: add raw (-bluestore) mode ([pr#32095](#), Sage Weil)
- ceph-volume: add sizing arguments to prepare ([pr#32235](#), Jan Fajerski)
- ceph-volume: add utility functions ([pr#27282](#), Mohamad Gebai)
- ceph-volume: allow raw block devices everywhere ([pr#31410](#), Jan Fajerski)
- ceph-volume: allow to skip restorecon calls ([pr#31421](#), Alfredo Deza)
- ceph-volume: api/lvm: check if list of LVs is empty ([pr#30101](#), Rishabh Dave)
- ceph-volume: assume msgrV1 for all branches containing mimic ([pr#31592](#), Jan Fajerski)
- ceph-volume: avoid calling zap\_lv with a LV-less VG ([pr#33283](#), Jan Fajerski)
- ceph-volume: batch bluestore fix create\_lvs call ([pr#32929](#), Jan Fajerski)
- ceph-volume: batch ensure device lists are disjoint ([pr#27754](#), Jan Fajerski)
- ceph-volume: check if we run in an selinux environment ([pr#31809](#), Jan Fajerski)
- ceph-volume: check if we run in an selinux environment, now also in py2 ([pr#31814](#), Jan Fajerski)
- ceph-volume: Dereference symlink in lvm list ([pr#32525](#), Benoxc3xaet Knecht)
- ceph-volume: detect ceph-disk osd if PARTLABEL is missing ([issue#40917](#), [pr#29401](#), Jan Fajerski)
- ceph-volume: do not fail when trying to remove crypt mapper ([pr#30490](#), Guillaume Abrioux)
- ceph-volume: dont keep device lists as sets ([pr#29683](#), Jan Fajerski)
- ceph-volume: dont remove vg twice when zapping filestore ([pr#33332](#), Jan Fajerski)
- ceph-volume: dont try to test lvm zap on simple tests ([pr#29659](#), Jan Fajerski)
- ceph-volume: finer grained availability notion in inventory ([pr#32634](#), Jan Fajerski)
- ceph-volume: fix batch functional tests, idempotent test must check sxe2x80xa6 ([pr#29684](#), Jan Fajerski)

- ceph-volume: fix device unittest, mock has\_bluestore\_label ([pr#32655](#), Jan Fajerski)
- ceph-volume: fix has\_bluestore\_label() function ([pr#33074](#), Guillaume Abrioux)
- ceph-volume: fix is\_ceph\_device for lvm batch ([pr#33223](#), Jan Fajerski, Dimitri Savineau)
- ceph-volume: fix lvm list ([pr#33077](#), Guillaume Abrioux)
- ceph-volume: fix regression and improve output in lvm list ([pr#33112](#), Jan Fajerski)
- ceph-volume: fix stderr failure to decode/encode when redirected ([pr#30274](#), Alfredo Deza)
- ceph-volume: fix the integer overflow ([pr#32106](#), dongdong tao)
- ceph-volume: fix warnings raised by pytest ([pr#30422](#), Rishabh Dave)
- ceph-volume: import mock.mock instead of unittest.mock (py2) ([pr#31816](#), Jan Fajerski)
- ceph-volume: look for rotational data in lsblk ([pr#26957](#), Andrew Schoen)
- ceph-volume: lvm: get\_device\_vgs() filter by provided prefix ([pr#33478](#), Jan Fajerski, Yehuda Sadeh)
- ceph-volume: make get\_devices fs location independent ([pr#31574](#), Jan Fajerski)
- ceph-volume: minor clean-up of simple scan subcommand help ([pr#31821](#), Michael Fritch)
- ceph-volume: minor optimizations related to class Volumess use ([pr#29665](#), Rishabh Dave)
- ceph-volume: mokeypatch calls to lvm related binaries ([pr#31197](#), Jan Fajerski)
- ceph-volume: never log to stdout, use stderr instead ([pr#29547](#), Jan Fajerski)
- ceph-volume: pass -ssh-config to pytest to resolve hosts when connecting ([issue#40063](#), [pr#28294](#), Alfredo Deza)
- ceph-volume: pass journal\_size as Size not string ([pr#33320](#), Jan Fajerski)
- ceph-volume: pre-install python-apt and its variants before test runs ([pr#30115](#), Alfredo Deza)
- ceph-volume: print most logging messages to stderr ([issue#38548](#), [pr#27675](#), Jan Fajerski)
- ceph-volume: PVolumes.filter shouldnt purge itself ([pr#30703](#), Rishabh Dave)

- ceph-volume: rearrange api/lvm.py ([pr#30867](#), Rishabh Dave)
- ceph-volume: refactor listing.py ([pr#31700](#), Rishabh Dave)
- ceph-volume: reject disks smaller than 5GB in inventory ([issue#40776](#), [pr#29041](#), Jan Fajerski)
- ceph-volume: revert -no-tmpfs change ([pr#30788](#), Sage Weil)
- ceph-volume: silence ceph-bluestore-tool failures ([pr#33371](#), Sxc3xa9bastien Han)
- ceph-volume: skip osd creation when already done ([pr#33086](#), Guillaume Abrioux)
- ceph-volume: strip \_dmcrypt suffix in simple scan json output ([pr#33079](#), Jan Fajerski)
- ceph-volume: systemd fix typo in log message ([pr#30497](#), Manu Zurmx3xbch1)
- ceph-volume: terminal: encode unicode when writing to stdout ([pr#27148](#), Alfredo Deza, Kefu Chai)
- ceph-volume: use centos8 for functional testing ([pr#33174](#), Jan Fajerski)
- ceph-volume: use correct extents if using db-devices and >1 osds\_per\_device ([pr#32177](#), Fabian Niepelt)
- ceph-volume: use fsync for dd command ([pr#31479](#), Rishabh Dave)
- ceph-volume: use get\_device\_vgs in has\_common\_vg ([pr#33246](#), Jan Fajerski)
- ceph-volume: use python3 compatible print ([pr#30790](#), Kyr Shatskyy)
- ceph-volume: use the Device.rotational property instead of sys\_api ([pr#28060](#), Andrew Schoen)
- ceph-volume: use the OSD identifier when reporting success ([pr#29762](#), Alfredo Deza)
- ceph-volume: util: look for executable in \$PATH ([pr#31787](#), Shyukri Shyukriev)
- ceph-volume: util: Use proper param substitution ([pr#28448](#), Shyukri Shyukriev)
- ceph-volume: VolumeGroups.filter shouldn't purge itself ([pr#30707](#), Rishabh Dave)
- ceph-volume: when testing disable the dashboard ([pr#29387](#), Andrew Schoen)
- ceph.in: disable ASAN if libasan is not found ([pr#28247](#), Kefu Chai)
- ceph.in: do not preload asan even if not needed ([pr#28703](#), Kefu Chai)
- ceph.in: do not preload libasan if it is found ([pr#28275](#), Kefu Chai)
- ceph.in: print decoded output in interactive mode ([pr#33099](#), Jun Su)

- cephadm: -cap-add=SYS\_PTRACE ([pr#33442](#), Sage Weil)
- cephadm: Add ability to deploy grafana container ([pr#32491](#), Paul Cuzner)
- cephadm: add ability to specify a timeout ([pr#32049](#), Michael Fritch)
- cephadm: add alertmanager deployment feature ([pr#32949](#), Sage Weil, Paul Cuzner)
- cephadm: add assert foo is not None for mypy check ([pr#33876](#), Kefu Chai)
- cephadm: add grafana adopt ([pr#33746](#), Eric Jackson)
- cephadm: add locking ([pr#32334](#), Sage Weil)
- cephadm: add nfs-ganesha deployment ([pr#33064](#), Michael Fritch)
- cephadm: add prepare-host ([pr#33374](#), Sage Weil)
- cephadm: add prometheus adopt ([pr#33438](#), Eric Jackson)
- cephadm: add reconfig service action ([pr#32281](#), Sage Weil)
- cephadm: add start/stop hooks and c-v activate on container start ([pr#32158](#), Sage Weil)
- cephadm: Add Zypper packager (openSUSE/SLES) ([pr#33461](#), Kristoffer Grxc3xb6nlund)
- cephadm: add -retry arg ([pr#33342](#), Michael Fritch)
- cephadm: add {add,rm}-repo commands ([pr#33062](#), Sage Weil)
- cephadm: add-repo: add -version ([pr#33961](#), Sage Weil)
- cephadm: adopt fixes ([pr#32995](#), Sage Weil)
- cephadm: allow multiple get\_parm() calls ([pr#33437](#), Sage Weil)
- cephadm: allow skipping prepare\_host in bootstrap step ([pr#33504](#), Kiefer Chang)
- cephadm: allow users to provide their dashboard cert during bootstrap ([pr#33472](#), Daniel-Pivonka)
- cephadm: also return JSON decode error ([pr#33433](#), Sebastian Wagner)
- cephadm: bootstrap: avoid repeat chars in generated password ([pr#32332](#), Sage Weil)
- cephadm: bootstrap: deploy monitoring stack by default ([pr#33936](#), Sage Weil)
- cephadm: bootstrap: nag about telemetry ([pr#33517](#), Sage Weil)
- cephadm: bootstrap: wait for mgr to restart after enabling a module ([pr#33857](#), Sage Weil)

- cephadm: bootstrap: warn on fqdn hostname ([pr#33042](#), Sage Weil)
- cephadm: check for both chrony service names ([pr#33369](#), Sage Weil)
- cephadm: check for both ntp.service and ntpd.service ([pr#32302](#), Sage Weil)
- cephadm: clean up the systemd unit and ceph-crash shutdown behavior ([pr#32685](#), Sage Weil)
- cephadm: correct ipv6 support in port open detection ([pr#32286](#), Paul Cuzner)
- cephadm: create /var/run/ceph/\$fsid as needed ([pr#32390](#), Sage Weil)
- cephadm: disable node-exporter cpu/memory limits for the time being ([pr#33133](#), Sage Weil)
- cephadm: drop sha256: prefix on container id ([pr#32300](#), Sage Weil)
- cephadm: error out on filestore OSDs ([pr#33395](#), Sage Weil)
- cephadm: fix adoption safety check ([pr#33445](#), Sage Weil)
- cephadm: fix ceph version probe ([pr#33136](#), Sage Weil)
- cephadm: fix container cleanup ([pr#32282](#), Sage Weil)
- cephadm: fix datetime regexp to capture at most 6 digits ([pr#33932](#), Michael Fritch)
- cephadm: fix deploy crash when no args.fsid ([pr#33248](#), Michael Fritch)
- cephadm: fix error handing in command\_check\_host() ([pr#33048](#), Guillaume Abrioux)
- cephadm: fix failure when getting keyring for deploying daemons ([pr#33679](#), Kiefer Chang)
- cephadm: fix help message for bootstrap -mgr-id ([pr#32640](#), Sage Weil)
- cephadm: fix inspect-image ([pr#33109](#), Sage Weil)
- cephadm: fix logging defaults ([pr#32641](#), Sage Weil)
- cephadm: fix name argument parsing during image check for non-ceph components ([pr#33114](#), Daniel-Pivonka)
- cephadm: Fix Py3 ConfigParser deprecation warnings ([pr#32218](#), Michael Fritch)
- cephadm: fix tox DeprecationWarning ([pr#32753](#), Michael Fritch)
- cephadm: fix v1/v2 ip/addrv handling; explicitly check bind to ip:port ([pr#32392](#), Sage Weil)
- cephadm: fix alertmanager not implemented yet ([pr#33694](#), Patrick Seidensal)

- cephadm: flag dashboard user to change password ([pr#32990](#), Daniel-Pivonka)
- cephadm: further simplify mon setup ([pr#33952](#), Sage Weil)
- cephadm: implement install command ([pr#33979](#), Sage Weil)
- cephadm: improve handling of crash agent container ([pr#33189](#), Sage Weil)
- cephadm: include daemon/unit id in unit name ([pr#32970](#), Sage Weil)
- cephadm: Infer ceph image ([pr#33829](#), Sage Weil, Ricardo Marques)
- cephadm: infer the fsid by name ([pr#32795](#), Michael Fritch)
- cephadm: KillMode=none in unit file ([pr#33162](#), Sage Weil)
- cephadm: leave backup when removing stateful daemons ([pr#33973](#), Sage Weil)
- cephadm: make add-repo -release and -version independent ([pr#34034](#), Sage Weil)
- cephadm: merge -config-and-keyring and -config-json args ([pr#33870](#), Michael Fritch)
- cephadm: misc upgrade fixes ([pr#32794](#), Sage Weil)
- cephadm: no -no-systemd arg to ceph-volume deactivate ([pr#32886](#), Sage Weil)
- cephadm: only infer image for shell, run, inspect-image, pull, ceph-volume ([pr#34030](#), Sage Weil)
- cephadm: podman inspect: image field was called ImageID ([pr#32616](#), Sebastian Wagner)
- cephadm: prepare-host: do not create Packager unless we need it ([pr#33443](#), Sage Weil)
- cephadm: pull: strip newline from version string ([pr#33446](#), Sage Weil)
- cephadm: python3 shebang ([pr#32378](#), Sage Weil)
- cephadm: re-introduce the podman logs command ([pr#33089](#), Michael Fritch)
- cephadm: Read ceph version from io.ceph.version label if set ([pr#32982](#), Kristoffer Grxc3xb6nlund)
- cephadm: Refactor, prepare for other adoptions ([pr#33672](#), Eric Jackson)
- cephadm: relabel /etc/ganesha mount ([pr#34098](#), Sage Weil)
- cephadm: remove orphan daemons ([pr#33830](#), Sage Weil)
- cephadm: remove logs command ([pr#32752](#), Michael Fritch)

- cephadm: Rename tox tests ceph-daemon -> cephadm ([pr#32353](#), Michael Fritch)
- cephadm: report image name for stopped daemons ([pr#33190](#), Sage Weil)
- cephadm: report version for grafana prom etc ([pr#33804](#), Sage Weil)
- cephadm: shell: allow -e ([pr#33191](#), Sage Weil)
- cephadm: shell: default to config and keyring in /etc/ceph, if present ([pr#33793](#), Sage Weil)
- cephadm: shell: do not bind ceph.conf twice ([pr#32425](#), Sage Weil)
- cephadm: shell: keep .bash\_history in /var/log/ceph/\$fsid ([pr#33519](#), Sage Weil)
- cephadm: show contextual message when port is in use ([pr#32560](#), Michael Fritch)
- cephadm: simplify Monitoring.components structure ([pr#32977](#), Michael Fritch)
- cephadm: SO\_REUSEADDR when doing bind check ([pr#32712](#), Sage Weil)
- cephadm: streamline bootstrap a bit ([pr#33980](#), Sage Weil)
- cephadm: support deployment of node-exporter ([pr#32340](#), Paul Cuzner)
- cephadm: support deployment of prometheus container ([pr#32198](#), Sebastian Wagner, Paul Cuzner)
- cephadm: switch grafana image to the ceph repo ([pr#34082](#), Paul Cuzner)
- cephadm: update unit.\\* atomically ([pr#33895](#), Sage Weil)
- cephadm: use appropriate default image for non-ceph components ([pr#33069](#), Sage Weil)
- cephadm: use spec to deploy crash on every host ([pr#33658](#), Sage Weil)
- cephadm: use sh instead of bash during enter ([pr#33822](#), Michael Fritch)
- cephadm: wait longer for things to come up ([pr#33216](#), Sage Weil)
- cephfs,common,core: global: disable THP for Ceph daemons ([pr#31582](#), Patrick Donnelly, Mark Nelson)
- cephfs,common,rbd: common/config\_proxy: hold lock while accessing mutable container ([pr#29809](#), Jason Dillaman)
- cephfs,common: common/secret.c: fix key parsing when doing a remount ([pr#28148](#), Luis Henriques)
- cephfs,common: osdc: should release the rwlock before waiting ([pr#29686](#), Kefu Chai)

- cephfs,core: mds/MDSDaemon: fix asok exit and respawn commands ([pr#32251](#), Sage Weil)
- cephfs,core: msg/async: perform the v2 resets in proper EventCenter ([pr#30717](#), Radoslaw Zarzynski)
- cephfs,core: qa/suites/rados/mgr/tasks/module\_selftest: whitelist mgr client getting backlisted ([issue#40867](#), [pr#29169](#), Sage Weil)
- cephfs,core: qa/suites/upgrade: a few more octopus fixes ([pr#32853](#), Sage Weil)
- cephfs,core: qa: log warning on scrub error ([pr#32739](#), Patrick Donnelly)
- cephfs,core: src/: define ceph\_release\_t and use it ([pr#27855](#), Kefu Chai)
- cephfs,mgr,mon: mon/MDSMonitor: enforce mds\_join\_fs cluster affinity ([pr#33194](#), Patrick Donnelly)
- cephfs,mgr,mon: mon/MgrMonitor: blacklist previous instance of ceph-mgr during failover ([pr#31797](#), Patrick Donnelly)
- cephfs,mgr,pybind: mgr/prometheus: export standby mds metadata ([pr#29996](#), lei01.liu)
- cephfs,mgr,pybind: mgr/volumes: minor enhancements and fixes ([issue#40429](#), [pr#28706](#), Ramana Raja)
- cephfs,mgr: mds/MDSRank: report state to mgr as mds id, not rank ([pr#31231](#), Patrick Donnelly, Sage Weil)
- cephfs,mgr: mgr/volume: ceph cephfs metadata pool pg\_num\_min and bias ([pr#27374](#), Sage Weil)
- cephfs,mgr: mgr/volumes: cleanup libcephfs handles on plugin shutdown ([issue#42299](#), [pr#30890](#), Venky Shankar)
- cephfs,mgr: pybind/mgr/volumes: use py3 items iterator ([pr#31986](#), Patrick Donnelly)
- cephfs,mgr: qa: use skipTest method instead of exception ([pr#27761](#), Patrick Donnelly)
- cephfs,mon: mon/MDSMonitor: cleanup check\_subs ([pr#32308](#), Patrick Donnelly)
- cephfs,mon: mon/MDSMonitor: handle standby already without fscid ([pr#32585](#), Patrick Donnelly)
- cephfs,pybind: libcephfs: add missing declaration of ceph\_getaddrs() ([pr#32629](#), Kefu Chai)
- cephfs,pybind: mgr/volumes: add ceph fs subvolumegroup getpath command ([issue#40617](#), [pr#29103](#), Ramana Raja)

- cephfs,pybind: mgr/volumes: set uid/gid of FS clients mount as 0/0 ([issue#40927](#), [pr#29355](#), Ramana Raja)
- cephfs,pybind: pybind/cephfs: add cephfs python API removexattr() ([pr#30641](#), bingyi zhang)
- cephfs,pybind: pybind/cephfs: Add listxattr ([pr#32804](#), Varsha Rao)
- cephfs,rbd,tests: qa/tasks: drop object inherit ([pr#29843](#), Jos Collin)
- cephfs,rbd: osdc: using decltype(auto) instead of trailing return type ([pr#29931](#), Yao Zongyou)
- cephfs,tests: cephfs-shell: teuthology tests ([issue#39526](#), [pr#27872](#), Milind Changire)
- cephfs,tests: mgr/volumes: fs subvolume resize command ([pr#30054](#), Jos Collin)
- cephfs,tests: qa/cephfs: add test for ACLs ([pr#29421](#), Rishabh Dave)
- cephfs,tests: qa/cephfs: change deps for xfstests-dev on centos8 ([pr#32524](#), Rishabh Dave)
- cephfs,tests: qa/cephfs: dont test kclient on RHEL 7 ([pr#32582](#), Rishabh Dave)
- cephfs,tests: qa/cephfs: update xfstests-dev deps for RHEL 8 ([pr#33427](#), Rishabh Dave)
- cephfs,tests: qa/suites/powercycle: install build deps for building xfstest ([pr#33874](#), Kefu Chai)
- cephfs,tests: qa/tasks/cephfs/fuse\_mount: use python3 ([pr#32339](#), Sage Weil)
- cephfs,tests: qa/tasks: add exception in do\_thrash() ([pr#29067](#), Jos Collin)
- cephfs,tests: qa/tasks: DaemonWatchdog Expansion ([issue#10369](#), [issue#11314](#), [pr#28378](#), Jos Collin)
- cephfs,tests: qa/tasks: Fix raises that doesnt re-raise ([pr#30201](#), Jos Collin)
- cephfs,tests: qa/tasks: fixed typo in the comment ([pr#29759](#), Jos Collin)
- cephfs,tests: qa/tasks: improvements in vstart\_runner.py and mount.py ([pr#27481](#), Rishabh Dave)
- cephfs,tests: qa/tasks: upgrade command arguments checks in vstart\_runner.py ([pr#28198](#), Rishabh Dave)
- cephfs,tests: qa/tests: reduce number of jobs for kcephfs ([pr#27328](#), Yuri Weinstein)
- cephfs,tests: qa/tests: reduced number of jobs for kcephfs ([pr#27165](#), Yuri Weinstein)

Weinstein)

- cephfs, tests: qa/vstart\_runner.py: make run()s interface same as teuthologys run ([pr#33263](#), Rishabh Dave)
- cephfs, tests: qa: note timeout in debug message ([pr#32162](#), Patrick Donnelly)
- cephfs, tests: qa: stop DaemonWatchdog for each cluster in daemon roles ([pr#29821](#), Patrick Donnelly)
- cephfs, tests: qa: test fs:upgrade when running upgrade suite ([pr#31206](#), Patrick Donnelly)
- cephfs, tests: test: define ALLPERMS if not yet ([pr#30726](#), Kefu Chai)
- cephfs, tests: test\_cephfs\_shell: fix test\_du\_works\_for\_hardlinks ([pr#32168](#), Rishabh Dave)
- cephfs, tests: test\_cephfs\_shell: initialize stderr for run\_cephfs\_shell\_cmd() ([pr#31626](#), Rishabh Dave)
- cephfs, tests: test\_sessionmap: use sudo\_write\_file() from teuthology.misc ([pr#29123](#), Rishabh Dave)
- cephfs, tools: cephfs-journal-tool: fix crash and usage ([pr#32452](#), Xiubo Li)
- cephfs, tools: mount.ceph: fix incorrect options parsing ([pr#33197](#), Xiubo Li)
- cephfs, tools: vstart.sh: highlight presence of stray conf ([pr#31403](#), Milind Changire)
- cephfs: client: more precise CEPH\_CLIENT\_CAPS\_PENDING\_CAPSNAP ([pr#28685](#), Yan, Zheng)
- cephfs: mds: change how mds revoke stale caps ([issue#17854](#), [pr#26737](#), Yan, Zheng, Rishabh Dave)
- cephfs: mds: fix corner case of replaying open sessions ([pr#28456](#), Yan, Zheng)
- cephfs: Add doc for deploying cephfs-nfs cluster using rook ([pr#30914](#), Varsha Rao)
- cephfs: Allow mount.ceph to get mount info from ceph configs and keyrings ([pr#29817](#), Jeff Layton)
- cephfs: avoid map client\_caps been inserted by mistake ([pr#29304](#), XiaoGuoDong2019)
- cephfs: ceph-mds: dump all info of ceph\_file\_layout, InodeStoreBase, frag\_infixe2x80xa6 ([pr#28874](#), simon gao)
- cephfs: ceph-mds: set ceph\_mds cpu affinity ([pr#31712](#), qilianghong)

- cephfs: cephfs pybind: added lseek() function to cephfs pybind ([pr#27688](#), Xiaowei Chu)
- cephfs: cephfs-shell: Add command for setxattr, getxattr and listxattr ([pr#32570](#), Varsha Rao)
- cephfs: cephfs-shell: Add error message for invalid ls commands ([pr#28652](#), Varsha Rao)
- cephfs: cephfs-shell: add quota management ([issue#39165](#), [pr#27483](#), Milind Changire)
- cephfs: cephfs-shell: add snapshot management ([issue#38681](#), [pr#27467](#), Milind Changire)
- cephfs: cephfs-shell: Add stat command ([pr#27753](#), Varsha Rao)
- cephfs: cephfs-shell: Add tox for testing with flake8 ([pr#28239](#), Varsha Rao)
- cephfs: cephfs-shell: better complain info, when deleting non-empty directory ([issue#40864](#), [pr#30341](#), Shen Hang)
- cephfs: cephfs-shell: Catch OSError exceptions in lcd ([issue#40243](#), [pr#28473](#), Varsha Rao)
- cephfs: cephfs-shell: cd with no args must change CWD to root ([issue#40476](#), [pr#28793](#), Rishabh Dave)
- cephfs: cephfs-shell: changes related to read\_ceph\_conf() ([pr#32347](#), Rishabh Dave)
- cephfs: cephfs-shell: changes to stderr and stdout messages ([pr#30365](#), Rishabh Dave)
- cephfs: cephfs-shell: Convert paths type from string to bytes ([pr#29552](#), Varsha Rao)
- cephfs: cephfs-shell: du should ignore non-directory files ([issue#40371](#), [pr#28560](#), Rishabh Dave, Varsha Rao)
- cephfs: cephfs-shell: Fix df command errors ([pr#27894](#), Varsha Rao)
- cephfs: cephfs-shell: Fix flake8 blank line and indentation error ([pr#29149](#), Varsha Rao)
- cephfs: cephfs-shell: Fix hidden files and directories list by ls command ([pr#27266](#), Varsha Rao)
- cephfs: cephfs-shell: Fix ll command errors ([issue#40244](#), [pr#28475](#), Varsha Rao)
- cephfs: cephfs-shell: Fix ls -l ([pr#32801](#), Kotresh HR)

- cephfs: cephfs-shell: Fix mkdir relative path error ([pr#27822](#), Varsha Rao)
- cephfs: cephfs-shell: Fix multiple flake8 errors ([pr#28080](#), Varsha Rao)
- cephfs: cephfs-shell: Fix multiple flake8 errors ([pr#28433](#), Varsha Rao)
- cephfs: cephfs-shell: Fix multiple flake8 errors ([pr#29374](#), Varsha Rao)
- cephfs: cephfs-shell: Fix onecmd TypeError ([pr#29554](#), Varsha Rao)
- cephfs: cephfs-shell: Fix print of error messages to stdout ([pr#28447](#), Varsha Rao)
- cephfs: cephfs-shell: Fix rmdir -p issues and add tests for rmdir ([pr#31633](#), Varsha Rao)
- cephfs: cephfs-shell: fix string decoding for ls command ([issue#39404](#), [pr#27716](#), Milind Changire)
- cephfs: cephfs-shell: Fix TypeError in poutput() ([pr#28906](#), Varsha Rao)
- cephfs: cephfs-shell: Fix typo for mounting ([pr#28718](#), Varsha Rao)
- cephfs: cephfs-shell: fix unnecessary usage of to\_bytes for file paths ([issue#40455](#), [pr#28663](#), Patrick Donnelly)
- cephfs: cephfs-shell: fix various tracebacks ([issue#38743](#), [issue#38739](#), [issue#38741](#), [issue#38740](#), [pr#27235](#), Milind Changire)
- cephfs: cephfs-shell: make compatible with cmd2 versions after 0.9.13 ([pr#30585](#), Rishabh Dave)
- cephfs: cephfs-shell: make every command set a return value on failure ([pr#32213](#), Rishabh Dave)
- cephfs: cephfs-shell: print helpful message when conf file is not found ([pr#31460](#), Rishabh Dave)
- cephfs: cephfs-shell: py version fixes ([issue#40418](#), [pr#28638](#), Patrick Donnelly)
- cephfs: cephfs-shell: read options from ceph.conf ([pr#29964](#), Rishabh Dave)
- cephfs: cephfs-shell: rearrange code for convenience ([pr#31629](#), Rishabh Dave)
- cephfs: cephfs-shell: Remove extra length argument passed to setattr() ([pr#30802](#), Varsha Rao)
- cephfs: cephfs-shell: Remove str object references to attribute decode ([pr#27345](#), Varsha Rao)
- cephfs: cephfs-shell: Remove undefined variable files in do\_rm() ([pr#28710](#), Varsha Rao)

- cephfs: cephfs-shell: return non-zero value on error ([pr#30657](#), Rishabh Dave)
- cephfs: cephfs-shell: rewrite help text for put and get commands ([pr#30297](#), Rishabh Dave)
- cephfs: cephfs-shell: Use colorama module instead of colorize ([pr#27427](#), Varsha Rao)
- cephfs: ceph\_volume\_client: convert string to bytes object ([issue#40369](#), [issue#40800](#), [pr#28557](#), Rishabh Dave)
- cephfs: ceph\_volume\_client: decode d\_name before using it ([issue#39406](#), [pr#28196](#), Rishabh Dave)
- cephfs: client: add client\_fs mount option support ([pr#33506](#), Xiubo Li)
- cephfs: client: Add is\_dir() check before changing directory ([pr#32637](#), Varsha Rao)
- cephfs: client: add procession of SEEK\_HOLE and SEEK\_DATA in lseek ([pr#30416](#), Shen Hang)
- cephfs: client: add stx\_btime and stx\_version in cephfs.pyx ([pr#30206](#), huanwen ren)
- cephfs: client: add warning when cap != in->auth\_cap ([pr#30402](#), Shen Hang)
- cephfs: client: avoid length overflow by calling the lseek function ([pr#29626](#), wenpengLi)
- cephfs: Client: bump ll\_ref from int32 to uint64\_t ([pr#29136](#), Xiaoxi CHEN)
- cephfs: client: directory size always is zero lead to is\_quota\_bytes\_approaching lose efficacy ([pr#26104](#), guoyong)
- cephfs: client: disallow changing fuse\_default\_permissions option at runtime ([pr#32315](#), Zhi Zhang)
- cephfs: client: dont report any vxattrs to listxattr ([pr#29339](#), Jeff Layton)
- cephfs: client: fix bad error handling in ll\_lookup\_inode ([issue#40085](#), [pr#28324](#), Jeff Layton)
- cephfs: client: fix bad error handling in lseek SEEK\_HOLE / SEEK\_DATA ([pr#33480](#), Jeff Layton)
- cephfs: client: fix dir.rctime and snap.btime vxattr values ([pr#28116](#), David Disseldorp)
- cephfs: client: fix fuse client hang because its bad session PipeConnection to mds ([issue#39305](#), [pr#27482](#), Guan yunfei)

- cephfs: client: fix lazyio\_synchronize() to update file size ([pr#29705](#), Sidharth Anupkrishnan)
- cephfs: client: Fixes for missing consts SEEK\_DATA and SEEK\_HOLE on alpine linux ([pr#33104](#), Stefan Bischoff)
- cephfs: client: nfs-ganesha with cephfs client, removing dir reports not empty ([issue#40746](#), [pr#29005](#), Peng Xie)
- cephfs: client: optimize rename operation under different quota root ([issue#39715](#), [pr#28077](#), Zhi Zhang)
- cephfs: client: remove Inode.dir\_contacts field and handle bad whence value to llseek gracefully ([pr#30580](#), Jeff Layton)
- cephfs: client: remove unused variable ([pr#31509](#), [su\\_nan@inspur.com](#))
- cephfs: client: return -EIO when sync file which unsafe reqs have been dropped ([issue#40877](#), [pr#29167](#), simon gao)
- cephfs: client: set snapdirs link count to 1 ([pr#28545](#), Yan, Zheng)
- cephfs: client: support the fallocate() when fuse version >= 2.9 ([issue#40615](#), [pr#28831](#), huanwen ren)
- cephfs: Client: unlink dentry for inode with llref=0 ([issue#40960](#), [pr#29321](#), Xiaoxi CHEN)
- cephfs: client: \_readdir\_cache\_cb() may use the readdir\_cache already clear ([issue#41148](#), [pr#29526](#), huanwen ren)
- cephfs: clientxefxbcx9aEINVAL may be returned when offset is 0 ([pr#30312](#), wenpengLi)
- cephfs: Deploy ganesha daemons with vstart ([pr#31527](#), Varsha Rao)
- cephfs: expose snapshot creation time as new ceph.snap.btime vxattr ([pr#27077](#), David Disseldorp)
- cephfs: include: fix interval\_set const\_iterator call operator type ([pr#32185](#), Patrick Donnelly)
- cephfs: libcephfs: Add Tests for LazyIO ([issue#40283](#), [pr#28834](#), Sidharth Anupkrishnan)
- cephfs: mds : clean up data written to unsafe inodes ([pr#30969](#), simon gao)
- cephfs: mds : optimization functions, get\_dirfrags\_under, to speed up processing directories with tens of millions of files ([pr#31123](#), simon gao)
- cephfs: mds,mon: deprecate CephFS inline\_data support ([pr#29824](#), Jeff Layton)

- cephfs: mds/client: inode number delegation ([pr#31817](#), Jeff Layton)
- cephfs: mds/FSMap: fix adjust\_standby\_fscid ([pr#32709](#), Sage Weil)
- cephfs: mds/OpenFileTable: match MAX\_ITEMS\_PER\_OBJ to osd\_deep\_scrub\_large\_omap\_object\_key\_threshold ([pr#31232](#), Vikhyat Umrao)
- cephfs: mds/server:mds: drop reconnect message from non-existent session ([issue#39026](#), [pr#27256](#), Shen Hang)
- cephfs: messages: make CephFS messages safe ([pr#31330](#), Patrick Donnelly)
- cephfs: mgr / volume: refactor [sub]volume ([issue#39969](#), [pr#28082](#), Venky Shankar)
- cephfs: mgr / volumes: background purge queue for subvolumes ([issue#40036](#), [pr#28003](#), Patrick Donnelly, Venky Shankar)
- cephfs: mgr/dashboard: CephFS class issues with strings ([pr#29353](#), Volker Theile)
- cephfs: mgr/volume: adapt arg passing to ServiceSpec ([pr#33687](#), Joshua Schmid)
- cephfs: mgr/volumes: add mypy support ([pr#33674](#), Michael Fritch)
- cephfs: mgr/volumes: check for string values in uid/gid ([pr#31961](#), Jos Collin)
- cephfs: mgr/volumes: cleanup leftovers from earlier purge job implementation ([pr#30886](#), Venky Shankar)
- cephfs: mgr/volumes: cleanup on fs create error ([pr#32459](#), Jos Collin)
- cephfs: mgr/volumes: clone from snapshot ([issue#24880](#), [pr#32030](#), Venky Shankar)
- cephfs: mgr/volumes: convert string to bytes object ([issue#39750](#), [pr#28380](#), Rishabh Dave)
- cephfs: mgr/volumes: drop unused size ([pr#30185](#), Jos Collin)
- cephfs: mgr/volumes: drop unused variable vol\_name ([pr#31780](#), Joshua Schmid)
- cephfs: mgr/volumes: fail removing subvolume with snapshots ([issue#43645](#), [pr#32696](#), Venky Shankar)
- cephfs: mgr/volumes: fetch trash and clone entries without blocking volume access ([issue#44207](#), [pr#33413](#), Venky Shankar)
- cephfs: mgr/volumes: fix error message ([issue#40014](#), [pr#28407](#), Ramana Raja)
- cephfs: mgr/volumes: fix incorrect snapshot path creation ([pr#30654](#), Ramana Raja)
- cephfs: mgr/volumes: fix placement default value ([pr#33476](#), Sage Weil)
- cephfs: mgr/volumes: fix subvolume creation with quota ([issue#40152](#), [pr#28384](#), Ramana Raja)

- cephfs: mgr/volumes: fs subvolume resize inf command ([pr#31157](#), Jos Collin)
- cephfs: mgr/volumes: handle exceptions in purge thread with retry ([issue#41218](#), [issue#41219](#), [pr#29735](#), Venky Shankar)
- cephfs: mgr/volumes: improve volume deletion process ([pr#31762](#), Joshua Schmid)
- cephfs: mgr/volumes: list FS subvolumes, subvolume groups, and their snapshots ([pr#30476](#), Jos Collin)
- cephfs: mgr/volumes: minor fixes ([pr#29760](#), Ramana Raja)
- cephfs: mgr/volumes: prevent negative subvolume size ([pr#30058](#), Jos Collin)
- cephfs: mgr/volumes: protection for fs volume rm command ([pr#30407](#), Jos Collin)
- cephfs: mgr/volumes: refactor dir handle cleanup ([pr#30887](#), Jos Collin)
- cephfs: mgr/volumes: remove stale subvolume module ([pr#32645](#), Venky Shankar)
- cephfs: mgr/volumes: return string type to ceph-manager ([pr#30451](#), Venky Shankar)
- cephfs: mgr/volumes: sync inode attributes for cloned subvolumes ([issue#43965](#), [pr#33120](#), Venky Shankar)
- cephfs: mgr/volumes: uid, gid for subvolume create and subvolumegroup create commands ([pr#30336](#), Jos Collin)
- cephfs: mgr/volumes: unregister job upon async threads exception ([issue#44293](#), [pr#33547](#), Venky Shankar)
- cephfs: mgr/volumes: versioned subvolume provisioning ([pr#31763](#), Venky Shankar)
- cephfs: mon,mds: map mds daemons to a particular fs ([pr#32015](#), Sage Weil)
- cephfs: mon/MDSMonitor: use stringstream instead of dout for mds repaired ([issue#40472](#), [pr#28683](#), Zhi Zhang)
- cephfs: mon/MDSMonitor: warn when creating fs with default EC data pool ([pr#31494](#), Patrick Donnelly)
- cephfs: mount.ceph.c: do not pass nofail to the kernel ([pr#26992](#), Kenneth Waegeman)
- cephfs: mount.ceph: give a hint message when no mds is up or cluster is laggy ([pr#32164](#), Xiubo Li)
- cephfs: mount.ceph: new mount option alias - translate fs= option to mds\_namespace= ([pr#33491](#), Xiubo Li)
- cephfs: mount.ceph: properly handle -o strictatime ([pr#29518](#), Jeff Layton)

- cephfs: mount.ceph: remove arbitrary limit on size of name= option ([pr#32706](#), Jeff Layton)
- cephfs: mount: fix the debug log when keyring getting secret failed ([pr#33499](#), Xiubo Li)
- cephfs: octopus: Add FS subvolume clone cancel ([issue#44208](#), [pr#34018](#), Venky Shankar)
- cephfs: osdc/objecter: Fix last\_sent in scientific format and add age to ops ([pr#29818](#), Varsha Rao)
- cephfs: propagate ll\_releasedir errors ([pr#32548](#), David Disseldorp)
- cephfs: pybind / cephfs: remove static typing in LibCephFS.chown ([issue#42923](#), [pr#31756](#), Venky Shankar)
- cephfs: pybind/cephfs: Modification to error message ([pr#28628](#), Varsha Rao)
- cephfs: pybind/mgr: add cephfs subvolumes module ([issue#39610](#), [pr#27594](#), Ramana Raja)
- cephfs: pybind/test\_volume\_client: print python version correctly ([issue#40184](#), [pr#28221](#), Lianne)
- cephfs: qa/cephfs: fix test\_evict\_client ([pr#28411](#), Yan, Zheng)
- cephfs: qa/cephfs: make filelock\_interrupt.py work with python3 ([pr#32741](#), Yan, Zheng)
- cephfs: qa/cephfs: test case for auto reconnect after blacklisted ([pr#31200](#), Yan, Zheng)
- cephfs: qa/suites/fs/multifs/tasks/failover.yaml: disable RECENT\_CRASH ([pr#29363](#), Sage Weil)
- cephfs: qa/suites/fs: mon\_thrash test for fs ([issue#17309](#), [pr#27073](#), Jos Collin)
- cephfs: qa/tasks/cephfs: os.write takes bytes, not str ([pr#32359](#), Sage Weil)
- cephfs: qa/tasks: add remaining tests for fs volume ([pr#31884](#), Jos Collin)
- cephfs: qa/tasks: Better handling of thrasher names and \_\_init\_\_ calls ([pr#31207](#), Jos Collin)
- cephfs: qa/tasks: check if fs mounted in umount\_wait ([pr#30553](#), Jos Collin)
- cephfs: qa/tasks: Fix AttributeError: cant set attribute ([pr#31428](#), Jos Collin)
- cephfs: qa/tasks: upgrade the check for -c sudo option in vstart\_runner.py ([issue#39385](#), [pr#28199](#), Rishabh Dave)

- cephfs: qa/vstart\_runner.py: add more options ([pr#29906](#), Rishabh Dave)
- cephfs: qa: add debugging failed osd-release setting ([pr#29715](#), Patrick Donnelly)
- cephfs: qa: add upgrade test for volume upgrade from legacy ([pr#33636](#), Patrick Donnelly)
- cephfs: qa: allow client mount to reset fully ([issue#42213](#), [pr#30986](#), Venky Shankar)
- cephfs: qa: avoid subtree rep in test\_version\_splitting ([pr#33078](#), Patrick Donnelly)
- cephfs: qa: build v5.4 kernel ([pr#32763](#), Patrick Donnelly)
- cephfs: qa: decouple session map test from simple msgr ([issue#38803](#), [pr#27415](#), Patrick Donnelly)
- cephfs: qa: define centos version for fs:verify ([pr#32535](#), Patrick Donnelly)
- cephfs: qa: detect RHEL8 for yum package installation ([pr#32507](#), Patrick Donnelly)
- cephfs: qa: do not check pg count for new data\_isolated volume ([pr#31095](#), Patrick Donnelly)
- cephfs: qa: fix malformed suite config ([pr#29431](#), Patrick Donnelly)
- cephfs: qa: fix output check to not be sensitive to debugging ([pr#32163](#), Patrick Donnelly)
- cephfs: qa: fix testing kernel branch link ([pr#32854](#), Patrick Donnelly)
- cephfs: qa: fix various py3 cephfs qa bugs ([pr#32467](#), Patrick Donnelly)
- cephfs: qa: fix various py3 cephfs qa bugs x2 ([pr#32533](#), Patrick Donnelly)
- cephfs: qa: fs Ignore ceph.dir.pin: No such attribute errors in getattr tests for old kernel client ([pr#27377](#), Sidharth Anupkrishnan)
- cephfs: qa: fs/upgrade test fixes and cephfs feature bit updates for Octopus/Nautilus ([issue#39078](#), [issue#39077](#), [issue#39020](#), [pr#27303](#), Patrick Donnelly)
- cephfs: qa: have kclient tests use new mount.ceph functionality ([pr#30462](#), Jeff Layton)
- cephfs: qa: ignore expected MDS\_CLIENT\_LATE\_RELEASE warning ([issue#40968](#), [pr#29338](#), Patrick Donnelly)
- cephfs: qa: ignore RECENT\_CRASH for multimds snapshot testing ([pr#29911](#), Patrick Donnelly)

- cephfs: qa: ignore slow ops for ffsb workunit ([pr#32668](#), Patrick Donnelly)
- cephfs: qa: ignore trimmed cache items for dead cache drop ([pr#32644](#), Patrick Donnelly)
- cephfs: qa: install some dependencies for xfstests ([pr#32478](#), Patrick Donnelly)
- cephfs: qa: only restart MDS between tests ([pr#32532](#), Patrick Donnelly)
- cephfs: qa: remove requirement on simple msgr ([issue#39079](#), [pr#27301](#), Patrick Donnelly)
- cephfs: qa: rename kcephfs distro overrides ([pr#32639](#), Patrick Donnelly)
- cephfs: qa: save MDS epoch barrier ([pr#32642](#), Patrick Donnelly)
- cephfs: qa: sleep briefly after resetting kclient ([pr#29388](#), Patrick Donnelly)
- cephfs: qa: specify random distros in multimds ([pr#33080](#), Patrick Donnelly)
- cephfs: qa: tolerate ECONNRESET errcode during logrotate ([issue#41800](#), [pr#30809](#), Venky Shankar)
- cephfs: qa: update kclient testing to RHEL 7.6 ([pr#26662](#), Patrick Donnelly)
- cephfs: qa: use -D\_GNU\_SOURCE when compiling fsync-tester.c ([pr#32480](#), Patrick Donnelly)
- cephfs: qa: use hard\_reset to reboot kclient ([issue#37681](#), [pr#28825](#), Patrick Donnelly)
- cephfs: qa: use mimic-0 upgrade process ([pr#27731](#), Patrick Donnelly)
- cephfs: qa: use small default pg count for CephFS pools ([pr#30816](#), Patrick Donnelly)
- cephfs: qa: wait for MDS to come back after removing it ([issue#40967](#), [pr#29336](#), Patrick Donnelly)
- cephfs: qa: whitelist Error recovering journal for cephfs-data-scan ([pr#30971](#), Yan, Zheng)
- cephfs: qa: whitelist T00\_FEW\_PGS during Mimic deploy ([pr#31063](#), Patrick Donnelly)
- cephfs: Resolve a memory leak in cephfs/Resetter.cc ([pr#29302](#), XiaoGuoDong2019)
- cephfs: src/common: fix help text for echo option of cephfs-shell ([pr#33285](#), Rishabh Dave)
- cephfs: stop: Cleanly umount cephFS volumes ([pr#32024](#), Kotresh HR)

- cephfs: test/{fs,cephfs}: Get libcephfs and cephfs to compile with FreeBSD ([pr#30505](#), Willem Jan Withagen)
- cephfs: test: extend fs subvolume test to cover new interfaces ([issue#39949](#), [pr#27856](#), Venky Shankar)
- cephfs: test: use distinct subvolume/group/snapshot names ([issue#42646](#), [pr#31418](#), Venky Shankar)
- cephfs: test\_volumes: fix \_verify\_clone\_attrs call ([pr#33788](#), Ramana Raja)
- cephfs: test\_volume\_client: declare only one default for python version ([issue#40460](#), [pr#28194](#), Rishabh Dave)
- cephfs: test\_volume\_client: fix test\_put\_object\_versioned() ([issue#39405](#), [issue#39510](#), [pr#28692](#), Rishabh Dave)
- cephfs: test\_volume\_client: simplify test\_get\_authorized\_ids() ([pr#28171](#), Rishabh Dave)
- cephfs: tools/cephfs: make cephfs-data-scan scan\_links fix dentrys first ([pr#31680](#), Yan, Zheng)
- cephfs: Trivial comment and cleanup fixes for cephfs ([pr#27199](#), Jeff Layton)
- cephfs: vstart: add an alias for cephfs-shell to vstart\_environment.sh ([pr#27437](#), Jeff Layton)
- cephfs: vstart: generate environment script suitable for sourcing ([pr#27198](#), Jeff Layton)
- cephfs: vstart\_runner: allow the use of it with kernel mounts ([pr#30463](#), Jeff Layton)
- ceph\_argparse: increment matchcnt on kwargs ([pr#33004](#), Matthew Oliver)
- check rdma configuration and fix some logic problem ([pr#28344](#), Changcheng Liu)
- client/Client : Fix sign compare compiler warning ([pr#30719](#), Prashant D)
- cls/queue: fix data corruption in urgent data ([pr#33686](#), Yuval Lifshitz)
- cmake: support parallel build for rocksd ([pr#31781](#), Deepika Upadhyay)
- cmake: add add\_tox\_test() ([pr#29446](#), Kefu Chai)
- cmake: add cython\_cephfs to vstart target ([pr#28876](#), Kefu Chai)
- cmake: Add dpdk numa support ([pr#31841](#), Chunsong Feng, Hu Ye)
- cmake: Allow cephfs and ceph-mds to be build when building on FreeBSD ([pr#30494](#), Willem Jan Withagen)

- cmake: avoid rebuilding extensions, and using python-config ([pr#28920](#), Kefu Chai)
- cmake: boost fixes for ARM 32 bit ([pr#25729](#), Daniel Glaser)
- cmake: bump libceph-common SO version for compliance ([pr#30976](#), Nathan Cutler)
- cmake: check for MAJOR.MINOR version of python3 ([pr#27383](#), Kefu Chai, Boris Ranto)
- cmake: check for unaligned access ([pr#28936](#), Kefu Chai)
- cmake: check version of librdkafka ([pr#32237](#), Kefu Chai)
- cmake: cleanups ([pr#28252](#), Kefu Chai)
- cmake: cleanups ([pr#33500](#), Kefu Chai)
- cmake: compile crimson-auth with crimson::cflags ([pr#33296](#), Kefu Chai)
- cmake: dashboard: enable frontend on arm64 ([pr#30958](#), Kefu Chai)
- cmake: define mgr\_cap\_obj library when WITH\_MGR=OFF ([pr#31326](#), Casey Bodley)
- cmake: detect librt for POSIX time functions ([pr#31543](#), Kefu Chai)
- cmake: detect linker support ([pr#30781](#), Kefu Chai)
- cmake: Do a debug build by default ([pr#30799](#), Brad Hubbard)
- cmake: do not assume \${CMAKE\_GENERATOR} == make ([pr#27089](#), Kefu Chai)
- cmake: do not include \${CMAKE\_SOURCE\_DIR}/src/fmt/include ([pr#31761](#), Kefu Chai)
- cmake: do not include global\_context.cc multiple times ([pr#32607](#), Kefu Chai)
- cmake: do not link against unused libs ([pr#33247](#), Kefu Chai)
- cmake: do not use CMP0074 unless it is supported ([pr#31958](#), Kefu Chai)
- cmake: do not use CMP0093 unless it is supported ([pr#31960](#), Kefu Chai)
- cmake: exclude unittest\_alloc\_aging from all ([pr#33466](#), Kefu Chai)
- cmake: Fix build against ncurses with separate libtinfo ([pr#27443](#), Lars Wendler)
- cmake: Fix unaligned check on big-endian systems ([pr#30362](#), Ulrich Weigand)
- cmake: fix WITH\_UBSAN ([pr#28725](#), Casey Bodley)
- cmake: Improve test for 16-byte atomic support on IBM Z ([pr#32802](#), Ulrich Weigand)
- cmake: let vstart depend on radosgwd ([pr#32564](#), Kefu Chai)

- cmake: link ceph-fuse against librbd (pr#31531, Yong Wang)
- cmake: move crimson-crush to crimson/ (pr#33481, Kefu Chai)
- cmake: one run\_tox.sh to rule them all (pr#29457, Kefu Chai)
- cmake: pass arguments to crimson tests (pr#30655, Kefu Chai)
- cmake: pmem/pmdk changes to cmake (pr#28802, Scott Peterson, Xiaoyan Li)
- cmake: remove cython 0.29s subinterpreter check during install (pr#27067, Tim Serong)
- cmake: Removed unittest\_alloc\_aging from make check (pr#33397, Adam Kupczyk)
- cmake: require CMake v3.10.2 (pr#29291, Kefu Chai)
- cmake: require RocksDB 5.14 or higher (pr#29930, Ilsoo Byun)
- cmake: revert librados\_tp.so version from 3 to 2 (issue#39291, pr#27593, Nathan Cutler)
- cmake: rewrite Findgenl to support components argument (pr#28460, Kefu Chai)
- cmake: s/brotli\_libs/brotli\_libs/ (pr#30374, Kefu Chai)
- cmake: selectively rewrite install rpath (pr#30028, Kefu Chai)
- cmake: set empty INSTALL\_RPATH on crypto shared libs (issue#40398, pr#28593, Nathan Cutler)
- cmake: set empty RPATH for some test executables (pr#29922, Nathan Cutler)
- cmake: set empty-string RPATH for ceph-osd (issue#40295, pr#28508, Nathan Cutler)
- cmake: should expose \${C-ARES\_BINARY\_DIR} from c-ares (pr#33256, Kefu Chai)
- cmake: silence messages when cppcheck/IWYU is not found (pr#32140, Kefu Chai)
- cmake: support Seastar\_DPDK=ON option (pr#31110, Kefu Chai)
- cmake: Test for 16-byte atomic support on IBM Z (pr#30638, Ulrich Weigand)
- cmake: update FindBoost.cmake (pr#29396, Willem Jan Withagen)
- cmake: update FindBoost.cmake for 1.71 (pr#31317, Willem Jan Withagen)
- cmake: Update pmdk version to 1.7 (pr#32693, Yin, Congmin)
- cmake: update SPDK to build with GCC-9 (pr#28507, Kefu Chai)
- cmake: use BUILD\_ALWAYS for rebuilding external project (pr#28984, Kefu Chai)
- cmake: use GNU linker on FreeBSD (pr#30621, Willem Jan Withagen)

- cmake: use latest FindPython\\*.cmake ([pr#29100](#), Kefu Chai)
- cmake: use python2 by default ([pr#29148](#), Kefu Chai)
- cmake: use StdFilesystem::filesystem instead of stdc++fs ([pr#27149](#), Willem Jan Withagen)
- cmake: workaround of false alarm from ubsan ([pr#27094](#), Kefu Chai)
- CMakeLists.txt: fix typo in error message ([pr#28795](#), Kefu Chai)
- codeowners: Add ceph2.py to @ceph/orchestrators ([pr#32131](#), Sebastian Wagner)
- common,core,mon: src/: drop cct from cmd\_getval() ([pr#33010](#), Kefu Chai)
- common,core: common, auth: use boost::spirit to parse ceph.conf, escape quotes in exported auths ([issue#22227](#), [pr#28634](#), Kefu Chai, Gu Zhongyan)
- common,core: common,mgr,osd: pass string\_view as name ([pr#33167](#), Kefu Chai)
- common,core: common,osd: add hash algorithms for dedup fingerprint ([pr#28254](#), Myoungwon Oh)
- common,core: include/cpp-btree: use the same type when allocate/deallocate ([pr#33638](#), Kefu Chai)
- common,core: message,mgr: drop MessageFactory and friends and use ref\_t<> in mgr ([pr#27592](#), Patrick Donnelly, Kefu Chai)
- common,core: Remove dependence on using namespace: Build of common through osdc/Objecter.cc ([pr#27255](#), Adam C. Emerson)
- common,mgr: vstart.sh: set prometheus port for each mgr ([pr#33698](#), Alfonso Martxc3xadnez)
- common,mon: common/options: make mon\_clean\_pg\_upmaps\_per\_chunk unsigned ([pr#28509](#), Kefu Chai)
- common,rbd: common/ceph\_context: avoid unnecessary wait during service thread shutdown ([pr#30947](#), Jason Dillaman)
- common,rgw: common/Formatter: escape printed buffer in XMLFormatter::dump\_format\_va() ([issue#38121](#), [pr#26220](#), ashitakasam)
- common,rgw: rgw/OutputDataSocket: actually discard data on full buffer ([issue#40178](#), [pr#28415](#), Matt Benjamin)
- common,tests: python-common: Add mypy testing ([pr#31071](#), Sebastian Wagner)
- common,tests: test/test\_mempool: test accounting for btree\_map ([pr#33621](#), Adam Kupczyk)

- common, tools: src/common: add rabin chunking for dedup ([pr#26730](#), Myoungwon Oh, Hsuan-Heng, Wu)
- common, tools: vstart.sh: enable creating multiple OSDs backed by spdk backend ([pr#27841](#), Richael Zhuang)
- common, tools: vstart.sh: enable nfs-ganesha mgmt. in dashboard ([pr#33691](#), Alfonso Martxc3xadnez)
- common/config\_values: set seastar logging level per that of ceph ([pr#28792](#), Kefu Chai)
- common/options: remove unused ms\_msgr2\\_{sign, encrypt}\_messages ([pr#31818](#), Ilya Dryomov)
- common: crimson/osd: add -mkkey support ([pr#28534](#), Kefu Chai)
- common: .gitignore: ignore /src/python-common/build ([pr#32967](#), Alfonso Martxc3xadnez)
- common: add -log-early command line option ([pr#27419](#), Sage Weil)
- common: add bool log\_to\_file option ([pr#27044](#), Sage Weil)
- common: add comment about pod memory requests/limits ([pr#29331](#), Patrick Donnelly)
- common: add iterator-based string splitter ([pr#33696](#), Casey Bodley)
- common: add ref header ([pr#29119](#), Patrick Donnelly)
- common: auth/cephx: always initialize local variables ([pr#31154](#), Kefu Chai)
- common: auth/krb: fix Kerberos compile error ([issue#39948](#), [pr#28113](#), huangjun)
- common: avoid use of size\_t in options ([pr#28277](#), James Page)
- common: blobhash.h: remove extra [[fallthrough]] ([pr#28270](#), Thomas Johnson)
- common: blobhash: do not use cast for unaligned access ([pr#28099](#), Kefu Chai)
- common: buffer, denc: more constness ([pr#27767](#), Kefu Chai)
- common: buffer,crypto,tools: extract digest methods out of bufferlist ([pr#28486](#), Kefu Chai)
- common: buffer.h: remove list::iterator\_impl::advance(size\_t) ([pr#28278](#), Kefu Chai)
- common: ceph.in: use sys.\_exit when we dont shut down ([pr#33950](#), Sage Weil)
- common: ceph\_argparse: put args from env before existing ones ([pr#33243](#), Kefu Chai)

- common: Clang requires a default constructor, but it can be empty ([issue#39561](#), [pr#27844](#), Willem Jan Withagen)
- common: clean up CLUSTER\_CREATE and CREATE options ([pr#31584](#), Sage Weil)
- common: common,crimson: fixes to compile with clang and libc++ ([pr#32485](#), Kefu Chai)
- common: common,crimson: supporting admin-socket commands ([pr#32174](#), Ronen Friedman, Kefu Chai)
- common: common,log: use ISO 8601 datetime format ([pr#27799](#), Sage Weil, Casey Bodley)
- common: common,os: address string truncated warnings from GCC-9 ([pr#28289](#), Kefu Chai)
- common: common/admin\_socket: Added printing of error message ([pr#33380](#), Adam Kupczyk)
- common: common/bl: carry the bufferlist::\_carriage over std::moves ([pr#32937](#), Radoslaw Zarzynski)
- common: common/bl: fix memory corruption in bufferlist::claim\_append() ([pr#32823](#), Radoslaw Zarzynski)
- common: common/bl: fix the dangling last\_p issue ([pr#32702](#), Radoslaw Zarzynski)
- common: common/bloom\_filter: Fix endian issues ([pr#30527](#), Ulrich Weigand)
- common: common/ceph\_time: tolerate mono time going backwards ([pr#33699](#), Sage Weil)
- common: common/config: cleanups ([pr#33362](#), Jianpeng Ma)
- common: common/config: fix lack of normalize\_key\_name() apply ([pr#33558](#), Igor Fedotov)
- common: common/config: Remove unused code ([pr#28940](#), Jianpeng Ma)
- common: common/Finisher: remove some lock acquisitions ([pr#29495](#), Igor Fedotov)
- common: common/options: change default erasure-code-profile to k=2 m=2 ([pr#27656](#), Sage Weil)
- common: common/pick\_address.cc: silence GCC warning ([pr#32025](#), Kefu Chai)
- common: common/secret.c: dont pass uninitialized stack data to the kernel ([pr#30675](#), Ilya Dryomov)
- common: common/thread: Fix race condition in make\_named\_thread ([pr#31057](#), Adam C. Emerson)

- common: common/util: use ifstream to read from /proc files ([pr#32630](#), Kefu Chai)
- common: common/WorkQueue: narrow ThreadPool::\_lock in func worker ([pr#22411](#), Jianpeng Ma)
- common: crimson, common: introduce ceph::atomic and apply it on bufferlist ([pr#32766](#), Radoslaw Zarzynski)
- common: crimson, common: RefCountedObj doesnt use atomics in Seastar builds ([pr#28085](#), Radoslaw Zarzynski)
- common: crimson/osd: implement readable/lease related methods ([pr#30639](#), Kefu Chai)
- common: crimson/osd: Message has non-null ref to SocketConnection now ([pr#30124](#), Radoslaw Zarzynski)
- common: crimson: cleanups ([pr#33797](#), Kefu Chai)
- common: crimson: cleanups for clang build ([pr#32605](#), Kefu Chai)
- common: Cycles: Add support for IBM Z ([pr#30874](#), Ulrich Weigand)
- common: default pg\_autoscale\_mode=on for new pools ([pr#30112](#), Sage Weil)
- common: default pg\_autoscale\_mode=on for new pools ([pr#30475](#), Sage Weil)
- common: denc: fix build error by calling global sprintf ([pr#27572](#), Changcheng Liu)
- common: denc: slightly optimize container\_base::bound\_encode ([pr#24636](#), Radoslaw Zarzynski, Kefu Chai)
- common: denc: support enums wider than 8 bits ([pr#33673](#), Casey Bodley)
- common: dmclock: pick up fix to replace uint ([pr#28829](#), Kefu Chai)
- common: drop sharing of buffer::raw outside bufferlist ([pr#32806](#), Radoslaw Zarzynski)
- common: encode for std::list<T> doesnt use bl::copy\_in() anymore ([pr#32785](#), Radoslaw Zarzynski)
- common: FIPS: audit and switch some memset & bzero users ([pr#31692](#), Radoslaw Zarzynski)
- common: Fix 44373 and make a couple cleanups in ceph::timer ([pr#33771](#), Adam C. Emerson)
- common: fix clang build failures, and clean up warnings ([pr#26701](#), Adam C. Emerson)

- common: fix clang compile errors from cython\_modules ([pr#33056](#), Mark Kogan)
- common: fix compat of strerror\_r ([pr#30279](#), luo.runbing)
- common: fix deadlocky inflight op visiting in OpTracker ([pr#32364](#), Radoslaw Zarzynski)
- common: fix missing <stdio.h> include ([pr#31209](#), Willem Jan Withagen)
- common: fix parse\_env nullptr deref ([pr#28159](#), Patrick Donnelly)
- common: Fix the error handling logic in get\_device\_id ([pr#30636](#), Difan Zhang)
- common: fix typo in rgw\_user\_max\_buckets option long description ([pr#31571](#), Alfonso Martxc3xadnez)
- common: give lockdeps group name to OpenSSLs mutexes ([issue#40698](#), [pr#28987](#), Radoslaw Zarzynski)
- common: global/global\_context: always add \0 after strncpy() ([pr#28365](#), Kefu Chai)
- common: global/global\_init: do first transport connection after setuid() ([pr#28012](#), Roman Penyaev)

- common: global/pidfile: pass string\_view instead of ConfigProxy to pidfile\_wrx2x80xa6 ([pr#27975](#), Kefu Chai)
- common: handle return value from read(2) ([pr#32192](#), Patrick Donnelly)
- common: include, common: make ceph::bufferlist 32 bytes long on x86 ([pr#32934](#), Radoslaw Zarzynski)
- common: include/buffer: add operator+=() for list::iterator ([pr#33003](#), Kefu Chai)
- common: include/cpp-btree: drop btree::dump() ([pr#32692](#), Kefu Chai)
- common: include/interval\_set: rename some types ([pr#32415](#), Kefu Chai)
- common: include: switch mempool.h to ceph::atomic ([pr#33034](#), Radoslaw Zarzynski)
- common: json: JSONDecoder::err inherits from std::runtime\_error ([pr#27957](#), Casey Bodley)
- common: make cluster\_network work ([pr#27811](#), Jianpeng Ma)
- common: messages: MOSDPGCreate2 doesnt assume using namespace std ([pr#28342](#), Radoslaw Zarzynski)
- common: messages: remove MNop ([pr#27585](#), Kefu Chai)
- common: mgr/test\_orchestrator: Add dummy data ([pr#32182](#), Sebastian Wagner, Volker Theile)
- common: move gen\_rand\_alphanumeric() helpers into common ([pr#31567](#), Casey Bodley)
- common: move xattr -> os/filestore/os\_xattr ([pr#32219](#), David Disseldorp)
- common: msg/Message: remove unused local variables ([pr#29155](#), Kefu Chai)
- common: msg/msg\_types: use inet\_ntop(3) to render IP addresses ([pr#26987](#), Sage Weil)
- common: no need to include ceph\_assert.h ([pr#28255](#), Kefu Chai)
- common: octopus ([pr#27009](#), Sage Weil)
- common: optimize check\_utf8 ([pr#27628](#), Yibo Cai)
- common: optimize encode\_utf8 ([pr#27807](#), Yibo Cai)
- common: OutputDataSocket retakes mutex on error path ([issue#40188](#), [pr#28431](#), Casey Bodley)
- common: preforker: remove useless code ([pr#31714](#), Xiubo Li)
- common: python-common: Add drive selection ([pr#31021](#), Sebastian Wagner)

- common: python-common: add py.typed (PEP 561) ([pr#33236](#), Sebastian Wagner)
- common: python-common: Add small Readme ([pr#30587](#), Sebastian Wagner)
- common: python-common: avoid using setup\_requires in setup.py ([pr#31222](#), Sebastian Wagner)
- common: python-common: enable lint in tox tests ([pr#31068](#), Kiefer Chang)
- common: python-common: Fix typo in device type ([pr#31758](#), Volker Theile)
- common: python-common: Make Drive Group filter by AND, instead of OR ([pr#33625](#), Sage Weil, Sebastian Wagner)
- common: python-common: Make DriveGroupSpec a sub type of ServiceSpec ([pr#33817](#), Sebastian Wagner)
- common: random: added a deduction guide to make using the function obxe2x80xa6 ([pr#30224](#), Jesse Williamson)
- common: remove dead code in {safe,mutable}\_item\_history ([pr#32698](#), Radoslaw Zarzynski)
- common: remove unused \_STR and STRINGIFY macro ([pr#29605](#), Yao Zongyou)
- common: rename image to container\_image ([pr#30800](#), Sage Weil)
- common: Revert Merge pull request #33673 from cbodley/wip-denc-enum ([pr#33832](#), Sage Weil)
- common: selinux: Allow ceph to setsched ([pr#33404](#), Brad Hubbard)
- common: skip interfaces starting with lo in find\_ipv{4,6}\_in\_subnet() ([pr#32420](#), Jiawei Li)
- common: sort best-matched command by req argument count ([issue#40292](#), [pr#28510](#), Chang Liu)
- common: src/: remove execute permissions on nine source files ([pr#28781](#), J. Eric Ivancich)
- common: start logging for non-global\_init users ([pr#27352](#), Sage Weil)
- common: systemd: Wait 5 seconds before attempting a restart of an OSD ([pr#31550](#), Wido den Hollander)
- common: use of malloc.h is deprecated ([pr#29397](#), Willem Jan Withagen)
- common: zstd: upgrade to v1.4.0 ([pr#28656](#), Dan van der Ster)
- core,mgr,tools: osd,tools: Balancer fixes without all of the calc\_pg\_upmaps() rewrites ([pr#31774](#), David Zafman)

- core,mgr: mgr/ActivePyModules: drop GIL to register/unregister clients ([pr#33464](#), Sage Weil)
- core,mgr: mgr/alerts: simple module to send health alerts ([pr#30738](#), Sage Weil)
- core,mgr: mgr/DaemonServer: warn when we reject reports ([pr#31471](#), Sage Weil)
- core,mgr: mgr/pg\_autoscaler: add pg\_autoscale\_bias pool property and apply it to pg\_num selection ([pr#27154](#), Sage Weil)
- core,mgr: mgr/prometheus: report per-pool pg states ([pr#32370](#), Aleksei Zakharov)
- core,mgr: mgr/telemetry: add report\_timestamp to sent reports ([pr#27571](#), Dan Mick)
- core,mgr: mgr/telemetry: catch exception during requests.put ([pr#33070](#), Sage Weil)
- core,mgr: mgr/telemetry: obscure entity\_name with a salt ([pr#29330](#), Sage Weil)
- core,mgr: osd,mon,mgr: report /dev/disk/by-path paths for devices ([pr#32261](#), Sage Weil)
- core,mon: mon,osd: use get\_req<> instead of static\_cast<>(get\_req()) ([pr#30023](#), Kefu Chai)
- core,mon: mon/AuthMonitor: fix initial creation of rotating keys ([issue#40634](#), [pr#28850](#), Sage Weil)
- core,mon: mon/MonClient: add proper SRV priority support ([pr#27126](#), Kefu Chai)
- core,mon: mon/Monitor.cc: fix condition that checks for unrecognized auth mode ([pr#30015](#), Neha Ojha)
- core,mon: mon/Monitor.cc: print min\_mon\_release correctly ([pr#27107](#), Neha Ojha)
- core,mon: mon/OSDMonitor: clean up removed\_snap keys ([pr#30518](#), Sage Weil)
- core,mon: mon/OSDMonitor: expand iec\_options for osd pool set ([pr#31196](#), Sage Weil)
- core,mon: mon/OSDMonitor: Use generic priority cache tuner for mon caches ([issue#40870](#), [pr#28227](#), Sridhar Seshasayee)
- core,pybind: pybind/ceph\_argparse: avoid int overflow ([pr#33101](#), Kefu Chai)
- core,pybind: pybind/rados: fix set\_omap() crash on py3 ([pr#29096](#), Sage Weil)
- core,pybind: pybind/rados: fixed Python3 string conversion issue on get\_fsid ([issue#38381](#), [pr#26514](#), Jason Dillaman)
- core,rbd: common/config: use string\_view for keys ([pr#27097](#), Kefu Chai)

- core, rbd: osd/OSDCap: rbd profile permits use of rbd\_info ([issue#39973](#), [pr#28253](#), songweibin)
- core, rbd: osd/PrimaryLogPG: do not append outdata to TMAPUP ops ([pr#30457](#), Jason Dillaman)
- core, rgw, tests: librados, test, rgw: cleanups to deprecate safe\_cb related functions ([pr#31045](#), Kefu Chai)
- core, tests: ceph\_test\_cls\_hello: set RETURNVEC on the expected EINVAL request ([pr#33708](#), Sage Weil)
- core, tests: ceph\_test\_rados\_api\\_{watch\_notify, misc}: tolerate some timeouts ([pr#34011](#), Sage Weil)
- core, tests: Improvements to standalone tests ([pr#27279](#), David Zafman)
- core, tests: kv\_store\_bench: fix teuthology\_tests() return value ([pr#30293](#), luo rixin)
- core, tests: mon.test: improve validation and add a test for osd pool create ([pr#30538](#), Kefu Chai)
- core, tests: qa/objectstore: test with reduced value of osd\_memory\_target ([pr#27083](#), Neha Ojha)
- core, tests: qa/standalone/ceph-helpers: more osd debug ([issue#40666](#), [pr#28867](#), Sage Weil)
- core, tests: qa/standalone/misc/ok-to-stop: improve test ([pr#32738](#), Sage Weil)
- core, tests: qa/standalone/mon/health-mute.sh: misc fixes ([pr#29744](#), Sage Weil)
- core, tests: qa/standalone/osd/osd-backfill-recovery-log.sh: fix TEST\_backfill\_log\\_[1, 2] ([pr#32851](#), Neha Ojha)
- core, tests: qa/standalone/scrub/osd-scrub-snaps: snapmapper omap is now m ([pr#29774](#), Sage Weil)
- core, tests: qa/standalone/scrub/osd-scrub-test: wait longer for update ([pr#33809](#), Sage Weil)
- core, tests: qa/suites/rados/multimon: whitelist SLOW\_OPS while thrashing mons ([pr#29121](#), Sage Weil)
- core, tests: qa/suites/rados/perf: run on ubuntu ([pr#32355](#), Sage Weil)
- core, tests: qa/suites/rados/rest: run restful test on el8 ([pr#32920](#), Sage Weil)
- core, tests: qa/suites/rados/singleton-bluestore/cephtool: whitelist MON\_DOWN ([pr#33645](#), Sage Weil)

- core, tests: qa/suites/rados/singleton/all/lost-unfound\\*: whitelist SLOW\_OPS ([pr#32958](#), Sage Weil)
- core, tests: qa/suites/rados/singleton/all/recovery-preemption: fix pg log length ([pr#32898](#), Sage Weil)
- core, tests: qa/suites/rados/singleton/all/thrash-eio: whitelist slow request ([pr#33497](#), Sage Weil, Sridhar Seshasayee)
- core, tests: qa/suites/rados/thrash-old-clients: exclude ceph-daemon on nautilus installs ([pr#30817](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: rejigger v1 vs v2 settings ([pr#27249](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: tolerate MON\_DOWN ([pr#30577](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: use cephadm ([pr#32377](#), Sage Weil)
- core, tests: qa/suites/rados/thrash: force normal pg log length with cache tiering ([issue#38358](#), [issue#24320](#), [pr#28658](#), Sage Weil)
- core, tests: qa/suites/rados/thrash: increase async and partial recovery test coverage ([pr#30699](#), Neha Ojha)
- core, tests: qa/suites/rados/valgrind-leaks: independently verify we detect leaks on mon, osd, mgr ([pr#32946](#), Sage Weil)
- core, tests: qa/suites/rados/verify/tasks/mon\_recovery: whitelist SLOW\_OPS ([pr#33644](#), Sage Weil)
- core, tests: qa/suites/rados/verify: debug monc = 20 ([pr#32968](#), Sage Weil)
- core, tests: qa/suites/rados/verify: debug\_ms = 1 ([pr#33871](#), Sage Weil)
- core, tests: qa/suites/rados: move cephadm\_orchestrator to e18 ([pr#32407](#), Sage Weil)
- core, tests: qa/suites/upgrade/mimic-x-singleton: suppress TOO\_FEW\_PGS warning ([pr#31054](#), Sage Weil)
- core, tests: qa/suites/upgrade: fix mimic-x-singleton ([pr#32719](#), Sage Weil)
- core, tests: qa/suites/upgrade: misc fixes for octopus ([pr#32750](#), Sage Weil, Josh Durgin)
- core, tests: qa/tasks/cbt: run stop-all.sh while shutting down ([pr#31171](#), Sage Weil)
- core, tests: qa/tasks/ceph: restart: stop osd, mark down, then start ([pr#30196](#), Sage Weil)

- core, tests: qa/tasks/ceph\_manager: add -log-early to raw\_cluster\_cmd ([pr#32989](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: enable ceph-objectstore-tool via cephadm ([pr#32411](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: fix ceph-objectstore-tool incantations ([pr#32701](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: fix chmod on log dir during pg export copy ([pr#32943](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: fix post-osd-kill pg peered check ([pr#32737](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: make is\_{clean,recovered,active\_or\_down} less racy ([pr#28969](#), Sage Weil)
- core, tests: qa/tasks/mon\_thrash: sync force requires some force flags ([pr#30361](#), Sage Weil)
- core, tests: qa/tasks/radosbench: fix usage of -0 ([pr#33744](#), Sage Weil)
- core, tests: qa/tasks/thrashosds-health: disable osd\_max\_markdown behavior ([pr#33601](#), Sage Weil)
- core, tests: qa/workunits/cephtool/test.sh: delete test\_erasure pool ([pr#33188](#), Sage Weil)
- core, tests: qa/workunits/rados/test\_crash.sh: suppress core files ([pr#32724](#), Sage Weil)
- core, tests: qa: add basic omap testing capability ([pr#29120](#), Neha Ojha)
- core, tests: remove ceph\_test\_rados\_watch\_notify ([pr#34044](#), Sage Weil)
- core, tests: test/CMakeLists: disable memstore make check test ([pr#33473](#), Sage Weil)
- core, tests: test/librados: dont release handler if set\_pg\_num failed ([pr#32112](#), huangjun)
- core, tests: test/osd/safe-to-destroy.sh: fix typo ([pr#27651](#), Sage Weil)
- core, tests: test/pybind/test\_rados.py: test test\_aio\_remove ([pr#31003](#), Zhang Jiao)
- core, tests: test/unittest\_lockdep: do not start extra threads ([pr#32772](#), Kefu Chai)
- core, tests: test: Bump sleep time for slower machines ([pr#29494](#), David Zafman)

- core, tests: test: Make sure that extra scheduled scrubs dont confuse test ([issue#40078](#), [pr#28302](#), David Zafman)
- core, tests: tests/osd: fix typo in unittest\_osdmap ([pr#29790](#), huangjun)
- core, tests: tools/rados: use num ops instead of num objs for tracking outstanding IO ([pr#29734](#), Albert H Chen)
- core, tests: unittest\_lockdep: avoid any threads for death test ([pr#32765](#), Sage Weil)
- core, tools: ceph-objectstore-tool cant remove head with bad snapset ([pr#29919](#), David Zafman)
- core, tools: ceph.in: check ceph-conf returncode ([pr#30695](#), Dimitri Savineau)
- core, tools: src/tools/ceph-dedup-tool: Fix chunk scru ([pr#28765](#), Myoungwon Oh)
- core: ceph.in: only preload asan library for Debug build ([pr#27190](#), Kefu Chai)
- core: osd/ClassHandler: cleanups ([pr#28363](#), Kefu Chai)
- core: osd: add hdd, ssd and hybrid variants for osd\_snap\_trim\_sleep ([pr#28772](#), Neha Ojha)
- core: osdc/Objecter: use unique\_ptr<OSDMap> for Objecter::osdmap ([issue#38403](#), [pr#28397](#), Kefu Chai)
- core: Add structures for tracking in progress operations ([pr#28395](#), Samuel Just)
- core: auth: treat mgr the same as mon when selecting auth mode ([pr#33226](#), Yehuda Sadeh)
- core: backfill\_toofull seen on cluster where the most full OSD is at 1% ([pr#29857](#), David Zafman)
- core: ceph,pybind/mgr: a few py3 fixes ([pr#32187](#), Sage Weil)
- core: ceph-objectstore-tool: better error message if pgid and object do not match ([pr#30501](#), Sage Weil)
- core: ceph.in: Fix name retval is not defined error ([pr#33516](#), Varsha Rao)
- core: ceph.in: improve control-c handling ([pr#33352](#), Sage Weil)
- core: ceph.in: only shut down rados on clean exit ([pr#33825](#), Sage Weil)
- core: client: fix FTBFS due to bl::iterator::advance() ([pr#33085](#), Radoslaw Zarzynski)
- core: cls\_hello: fix typo ([pr#32976](#), Sage Weil)

- core: common,mon,osd: unify ceph tell and ceph daemon command sets ([pr#30217](#), Sage Weil)
- core: common,tools,crush,test: misc converity & klocwork fixes ([pr#29316](#), songweibin)
- core: common/admin\_socket: Increase socket timeouts ([pr#31623](#), Brad Hubbard)
- core: common/assert: include ceph\_abort\_msg(arg) arg in log output ([pr#27732](#), Sage Weil)
- core: common/blkdev: fix some problems with smart scraping ([pr#28848](#), Sage Weil)
- core: common/blkdev: get\_device\_id: behave if model is lvm and id\_model\_enc isnt there ([pr#27156](#), Sage Weil)
- core: common/blkdev: handle devices with ID\_MODEL as LVM PV ... but valid ID\_MODEL\_ENC ([pr#27020](#), Sage Weil)
- core: common/condition\_variable\_debug: do not assert() if sloppy ([pr#29854](#), Kefu Chai)
- core: common/config: behave when both POD\_MEMORY\_REQUEST and POD\_MEMORY\_LIMIT are set ([pr#29511](#), Sage Weil)
- core: common/config: less noise about configs from mon we cant apply ([pr#31988](#), Sage Weil)
- core: common/config: parse -default-\$option as a default value ([pr#27169](#), Sage Weil)
- core: common/config: update values when they are removed via mon ([pr#32091](#), Sage Weil)
- core: common/kv/rocksdb: Fixed async compations ([pr#26786](#), Adam Kupczyk)
- core: common/options.cc: Lower the default value of osd\_deep\_scrub\_large\_omap\_object\_key\_threshold ([pr#28782](#), Neha Ojha)
- core: common/options.cc: make rocksdb\_delete\_range\_threshold very high ([pr#33439](#), Neha Ojha)
- core: common/options: allow osd\_pool\_default\_pg\_autoscale\_mode to update a runtime ([pr#27821](#), Sage Weil)
- core: common/options: annotate some options; enable some runtime updates ([pr#27655](#), Sage Weil)
- core: common/options: decrease the default max\_omap\_entries\_per\_request ([pr#31506](#), Yan Jun)
- core: common/options: make secure mode non-experimental, and prefer/require it

- for mons ([pr#27012](#), Sage Weil)
- core: common/options: update mon\_crush\_min\_required\_version=hammer ([pr#27568](#), Sage Weil)
- core: common/PriorityCache: fix over-aggressive assert when mem limited ([pr#27763](#), Mark Nelson)
- core: common/PriorityCache: Implement a Cache Manager ([pr#27381](#), Mark Nelson)
- core: common/TextTable,mgr: standardize on 2 spaces between table columns ([pr#33138](#), Sage Weil)
- core: common/util: handle long lines in /proc/cpuinfo ([issue#38296](#), [pr#27707](#), Sage Weil)
- core: compressor/lz4: work around bug in liblz4 versions <1.8.2 ([pr#33584](#), Sage Weil, Dan van der Ster)
- core: crimson, osd: add support for Ceph Classes, part 1 ([pr#29651](#), Radoslaw Zarzynski)
- core: crimson/osd: add osd to crush when it boots ([pr#28689](#), Kefu Chai)
- core: crush/CrushCompiler: Fix \_\_replacement\_assert ([issue#39174](#), [pr#27506](#), Brad Hubbard)
- core: crush/CrushWrapper.cc: Fix sign compare compiler warning ([pr#31184](#), Prashant D)
- core: crush/CrushWrapper: behave with empty weight vector ([pr#32673](#), Kefu Chai)
- core: dencoder: include some missed types ([pr#27804](#), Greg Farnum)
- core: dmclock server side refactor ([pr#30650](#), Samuel Just)
- core: examples/librados: fix bufferlist::copy() in hello\_world.cc ([pr#33075](#), Radoslaw Zarzynski)
- core: Extract peering logic into a module for use in crimson ([pr#27874](#), Samuel Just, [sjust@redhat.com](mailto:sjust@redhat.com))
- core: feature: Health warnings on long network ping times, add dump\_osd\_network to get a report ([issue#40640](#), [pr#28755](#), David Zafman)
- core: Feature: Improvements to auto repair ([issue#38616](#), [pr#26942](#), David Zafman)
- core: global: ensure CEPH\_ARGS is decoded before early arg processing ([pr#32830](#), Jason Dillaman)
- core: global: explicitly call out EIO events in crash dumps ([pr#27386](#), Sage Weil)

- core: include/os: Make ceph\_le member private ([pr#30526](#), Ulrich Weigand)
- core: include/ceph\_features: fix typo ([pr#27353](#), Sage Weil)
- core: include/cpp-btree: cleanups ([pr#32443](#), Kefu Chai)
- core: init-ceph: wait longer before resending \$signal ([pr#27308](#), Kefu Chai)
- core: kv/KeyValueDB: fix estimate\_prefix\_size() ([pr#29842](#), Adam Kupczyk)
- core: kv/RocksDBStore: Add minimum key limit before invoking DeleteRange ([pr#31442](#), Mark Nelson)
- core: kv/RocksDBStore: make option: compaction\_threads/disableWAL/flusher\_txe2x80xa6 ([pr#32453](#), Jianpeng Ma)
- core: kv/RocksDBStore: tell rocksdb to set mode to 0600, not 0644 ([pr#30679](#), Sage Weil)
- core: kv: fix shutdown vs async compaction ([pr#32619](#), Sage Weil)
- core: kv: make delete range optional on number of keys ([pr#27317](#), Zengran Zhang)
- core: librados,osd,mon: remove traces of CEPH\_OSDMAP\_FULL ([pr#30614](#), Kefu Chai)
- core: Make dumping of reservation info congruent between scrub and recovery ([pr#30192](#), David Zafman)
- core: messages,osd: remove MPGStats::had\_map\_for ([pr#27026](#), Kefu Chai)
- core: messages: #include necessary header ([pr#27590](#), Kefu Chai)
- core: mgr/balancer: sort pool names in balancer ls output ([pr#32424](#), Sage Weil)
- core: mgr/balancer: tolerate pgs outside of target weight map ([pr#34014](#), Sage Weil)
- core: mgr/cephadm: health alert for stray services or hosts ([pr#32754](#), Sage Weil)
- core: mgr/crash: behave when posted crash has no backtrace ([pr#31643](#), Sage Weil)
- core: mgr/crash: raise warning about recent crashes and other improvements ([pr#29034](#), Sage Weil)
- core: mgr/DaemonServer: fix osd ok-to-stop for EC pools ([pr#32046](#), Sage Weil)
- core: mgr/DaemonServer: fix pg merge checks ([pr#34067](#), Sage Weil)
- core: mgr/DaemonServer: prevent pgp\_num reductions from outpacing pg\_num merges ([issue#38786](#), [pr#27473](#), Sage Weil)
- core: mgr/devicehealth: fix telemetry stops sending device reports after 48xe2x80xa6 ([pr#32903](#), Yaarit Hatuka)

- core: mgr/diskprediction\_cloud: Service unavailable ([issue#40478](#), [pr#28687](#), Rick Chen)
- core: mgr/diskprediction\_local: import scipy early to fix self-test deadlock ([pr#32102](#), Sage Weil)
- core: mgr/diskprediction\_local: some debug output during predict (and self-test) ([pr#31572](#), Sage Weil)
- core: mgr/MgrClient: fix open condition ([pr#31256](#), Sage Weil)
- core: mgr/MgrClient: fix open condition fix ([pr#31422](#), Sage Weil)
- core: mgr/MgrClient: fix tell mgr.x ... ([pr#31989](#), Sage Weil)
- core: mgr/pg\_autoscaler: complete event if pool disappears ([pr#30819](#), Sage Weil)
- core: mgr/pg\_autoscaler: default to pg\_num[\_min] = 16 ([pr#31636](#), Sage Weil)
- core: mgr/pg\_autoscaler: default to pg\_num[\_min] = 32 ([pr#32788](#), Neha Ojha)
- core: mgr/pg\_autoscaler: fix division by zero ([pr#33402](#), Sage Weil)
- core: mgr/pg\_autoscaler: only generate target\\_\\* health warnings if targets set ([pr#31638](#), Sage Weil)
- core: mgr/progress: behave if pgs disappear (due to a racing pg merge) ([issue#38157](#), [pr#27546](#), Sage Weil)
- core: mgr/progress: fix duration strings ([pr#34045](#), Sage Weil)
- core: mgr/progress: progress clear command should clear events in ceph -s ([pr#33400](#), Sage Weil)
- core: mgr/telemetry: add some more telemetry ([pr#31226](#), Sage Weil)
- core: mgr/telemetry: include pg\_autoscaler and balancer status ([pr#30871](#), Sage Weil)
- core: mgr/telemetry: send device telemetry via per-host POST to device endpoint ([pr#31225](#), Sage Weil)
- core: mgr/telemetry: split entity\_name only once (handle ids with dots) ([pr#33094](#), Dan Mick)
- core: Miscellaneous lost fixes ([pr#27599](#), Xinze Chi, Greg Farnum, linbing, shangfufei)
- core: mon, osd: parallel clean\_pg\_upmaps ([issue#40104](#), [pr#28373](#), xie xingguo)
- core: mon,msg/async: fix mon to mon authentication ([pr#27823](#), Sage Weil)

- core: mon,osd: add dead\_epoch, -dead flag to osd down ([pr#29221](#), Sage Weil)
- core: mon,osd: add no{out,down,in,out} flags on CRUSH nodes ([pr#27563](#), Sage Weil)
- core: mon,osd: deprecate forward and readforward cache modes ([pr#28944](#), Sage Weil)
- core: mon,osd: track history and past\_intervals for creating pgs ([pr#27696](#), Sage Weil)
- core: mon,osd: various octopus feature bits ([pr#27141](#), Sage Weil)
- core: mon/ConfigMap: search nested sections ([pr#31327](#), Sage Weil)
- core: mon/ConfigMonitor: fix handling of NO\_MON\_UPDATE settings ([pr#32726](#), Sage Weil)
- core: mon/ConfigMonitor: only propose if leader ([pr#32975](#), Sage Weil)
- core: mon/ConfigMonitor: prefix all global config options with global/ ([pr#32786](#), Sage Weil)
- core: mon/LogMonitor: add mon\_cluster\_log\_to\_file bool option ([pr#27343](#), Sage Weil)
- core: mon/MgrMonitor: fix null deref when invalid formatter is specified ([pr#29089](#), Sage Weil)
- core: mon/MgrMonitor: make mgr fail work with no arguments ([pr#33997](#), Sage Weil)
- core: mon/MgrStatMonitor: ensure only one copy of initial service map ([issue#38839](#), [pr#27101](#), Sage Weil)
- core: mon/MonClient: do not dereference auth\_supported.end() ([pr#27196](#), Kefu Chai)
- core: mon/MonClient: ENXIO when sending command to down mon ([pr#29090](#), Sage Weil, Greg Farnum)
- core: mon/MonClient: send logs to mon on separate schedule than pings ([pr#33732](#), Sage Weil)
- core: mon/MonClient: skip CEPHX\_V2 challenge if client doesnt support it ([pr#30523](#), Sage Weil)
- core: mon/Monitor: fail forwarded tell commands ([pr#33542](#), Sage Weil)
- core: mon/MonMap: encode (more) valid compat monmap when we have v2-only addrs ([pr#31472](#), Sage Weil)
- core: mon/MonmapMonitor: clean up empty created stamp in monmap ([issue#39085](#), [pr#27327](#), Sage Weil)

- core: mon/OSDMonitor.cc: Add current numbers of objects and bytes ([pr#18694](#), Shinobu Kinjo)
- core: mon/OSDMonitor.cc: better error message about min\_size ([pr#29184](#), Neha Ojha)
- core: mon/OSDMonitor: accept autoscale\_mode argument to osd pool create ([pr#33092](#), Sage Weil)
- core: mon/OSDMonitor: add check for crush rule size in pool set size command ([pr#30723](#), Vikhyat Umrao)
- core: mon/OSDMonitor: allow osd pool set pgp\_num\_actual ([pr#27010](#), Sage Weil)
- core: mon/OSDMonitor: allow pg\_num to increase when require\_osd\_release < N ([issue#39570](#), [pr#27928](#), Sage Weil)
- core: mon/OSDMonitor: Dont update mon cache settings if rocksdb is not used ([pr#32473](#), Sridhar Seshasayee)
- core: mon/OSDMonitor: fix format error ceph osd stat -format json ([pr#31399](#), Zheng Yin)
- core: mon/OSDMonitor: make memory autotune disable itself if no rocksdb ([pr#32044](#), Sage Weil)
- core: mon/OSDMonitor: tolerate duplicate MRemoveSaps messages ([issue#40774](#), [pr#29051](#), Sage Weil)
- core: mon/PGMap.h: disable network stats in dump\_osd\_stats ([pr#32406](#), Neha Ojha, David Zafman)
- core: mon/PGMap: drop indentation on df human output ([pr#30848](#), Sage Weil)
- core: mon/PGMap: fix summary display of >32bit pg states ([pr#33137](#), Sage Weil)
- core: mon/PGMap: use NONE for pg ls[-\\*] output too ([pr#32048](#), Sage Weil)
- core: mon/Session: only index osd ids >= 0 ([pr#32764](#), Sage Weil)
- core: More PeeringState and related cleanups to ease use in crimson ([pr#28048](#), Samuel Just)
- core: msg,auth: migrate msg/async V1 implementation to new Auth{Server,Client} interfaces ([pr#27566](#), Sage Weil)
- core: msg/async/frames\_v2.h: fix warning ([pr#27464](#), Sage Weil)
- core: msg/async/ProtocolV2: fix typo in register\_lossy\_clients fix ([pr#33559](#), Sage Weil)
- core: msg/async/ProtocolV[12]: add ms\_learn\_addr\_from\_peer ([pr#27341](#), Sage Weil)

- core: msg/async: clear\_payload when requeue\_sent ([pr#30211](#), Jianpeng Ma)
- core: msg/async: optimizations ([pr#26531](#), Jianpeng Ma)
- core: msg/auth: handle decode errors instead of throwing exceptions ([pr#31052](#), Sage Weil)
- core: msg/DispatchQueue: Set throttle stamp for local\_delivery ([pr#31137](#), Brad Hubbard)
- core: msg/Policy: limit unregistered anon connections to mon ([pr#33163](#), Sage Weil)
- core: msg/Policy: make stateless\_server default to anon (again) ([pr#33633](#), Sage Weil)
- core: objclass, osd: clean up the cls-host interface. Turn ClassHandler into singleton ([pr#29322](#), Radoslaw Zarzynski)
- core: object\_stat\_sum\_t decode broken if given older version ([issue#39284](#), [issue#39281](#), [pr#27564](#), David Zafman)
- core: os, osd: readv ([pr#30061](#), xie xingguo)
- core: os/bluestore: Add config observer for osd memory specific options ([pr#29606](#), Sridhar Seshasayee)
- core: os/filestore: assure sufficient leaves in pre-split ([issue#39390](#), [pr#27689](#), Jeegn Chen)
- core: os/Transaction: dump alloc hint flags in op ([pr#28881](#), Zengran Zhang)
- core: os: remove KineticStore ([pr#30653](#), Kefu Chai)
- core: osd,crimson: use make\_message for creating message ([pr#30412](#), Kefu Chai)
- core: osd,messages: changes for preparing for crimson-osd ([pr#27003](#), Kefu Chai)
- core: osd,mon: remove pg\_pool\_t::removed\_snaps ([pr#28330](#), Sage Weil)
- core: osd/ECTransaction,ReplicatedBackend: create op is new in octopus ([pr#29092](#), Sage Weil)
- core: osd/MissingLoc, PeeringState: remove osd from missing loc in purge\_strays() ([pr#30119](#), Neha Ojha)
- core: osd/MissingLoc.cc: do not rely on missing\_loc\_sources only ([pr#30226](#), Neha Ojha)
- core: osd/OSD.cc: make osd bench description consistent with parameters ([issue#39006](#), [pr#27600](#), Neha Ojha)

- core: osd/osd: add an err log to set numa affinity ([pr#30870](#), luo rixin)
- core: osd/OSD: auto mark heartbeat sessions as stale and tear them down ([issue#40586](#), [pr#28752](#), xie xingguo)
- core: osd/OSD: choose more heartbeat peers from different subtrees ([pr#33037](#), xie xingguo)
- core: osd/OSD: enhance osd numa affinity compatibility ([pr#31274](#), Dai zhiwei)
- core: osd/OSD: keep synchronizing with mon if stuck at booting ([pr#28404](#), xie xingguo)
- core: osd/OSD: Log slow ops/types to cluster logs ([pr#33328](#), Sridhar Seshasayee)
- core: osd/OSD: only wake up empty pqueue ([pr#28832](#), Jianpeng Ma)
- core: osd/OSD: prevent down osds from immediately rejoining the cluster ([pr#33039](#), xie xingguo)
- core: osd/osd: Refactor get\_iface\_numa\_node ([pr#31965](#), Dai zhiwei, luo rixin)
- core: osd/OSD: remove unused func enqueue\_peering\_evt\_front ([pr#32496](#), Jianpeng Ma)
- core: osd/OSD: remove unused parameter osdmap\_lock\_name ([pr#32514](#), Jianpeng Ma)
- core: osd/OSDCap: Check for empty namespace ([issue#40835](#), [pr#29146](#), Brad Hubbard)
- core: osd/OSDMap.cc: add more info in json output of osd stat ([pr#30344](#), Shen Hang)
- core: osd/OSDMap.cc: dont output over/underfull messages to lderr ([pr#31542](#), Neha Ojha)
- core: osd/OSDMap: add zone to default crush map ([pr#27070](#), Sage Weil)
- core: osd/OSDMap: calc\_pg\_upmaps - restrict optimization to origin pools only ([issue#38897](#), [pr#27142](#), xie xingguo)
- core: osd/OSDMap: consider overfull osds only when trying to do upmap ([pr#32368](#), xie xingguo)
- core: osd/OSDMap: do not trust partially simplified pg\_upmap\_item ([pr#30576](#), xie xingguo)
- core: osd/OSDMap: fix calc\_pg\_role ([pr#32132](#), Sage Weil)
- core: osd/OSDMap: health alert for non-power-of-two pg\_num ([pr#30525](#), Sage Weil)
- core: osd/OSDMap: Replace get\_out\_osds with get\_out\_existing\_osds ([issue#39154](#), [pr#27663](#), Brad Hubbard)

- core: osd/OSDMap: Show health warning if a pool is configured with size 1 ([pr#31416](#), Sridhar Seshasayee)
- core: osd/OSDMap: stop encoding osd\_state with >8 bits wide states only for old client ([pr#33814](#), xie xingguo)
- core: osd/osd\_types: bump up some encoding versions ([pr#29923](#), xie xingguo)
- core: osd/osd\_types: drop last\_backfill\_bitwise member ([pr#28766](#), Sage Weil)
- core: osd/osd\_types: fix {omap,hitset\_bytes}\_stats\_invalid handling on split/merge ([pr#30479](#), Sage Weil)
- core: osd/osd\_types: inc-recovery - add special handler for lost\_revert ([pr#29893](#), xie xingguo)
- core: osd/osd\_types: pool\_stat\_t::dump - fix num\_store\_stats field ([issue#39340](#), [pr#27633](#), xie xingguo)
- core: osd/PeeringState.cc: dont let num\_objects become negative ([pr#32305](#), Neha Ojha)
- core: osd/PeeringState.cc: skip peer\_purged when discovering all missing ([pr#32195](#), Neha Ojha)
- core: osd/PeeringState.h: Fix pg stuck in WaitActingChange ([pr#29669](#), chen qizhang)
- core: osd/PeeringState.h: get\_num\_missing() should report num\_missing() ([pr#30414](#), Neha Ojha)
- core: osd/PeeringState.h: ignore RemoteBackfillReserved in WaitLocalBackfillReserved ([pr#33525](#), Neha Ojha)
- core: osd/PeeringState: base lease support checks on features, not require\_osd\_release ([pr#30721](#), Sage Weil)
- core: osd/PeeringState: clear LAGGY and WAIT states on exiting Started ([pr#31864](#), Sage Weil)
- core: osd/PeeringState: disable read lease until require\_osd\_release >= octopus ([pr#30692](#), Sage Weil)
- core: osd/PeeringState: do not complain about past\_intervals constrained by oldest epoch ([pr#29747](#), Sage Weil)
- core: osd/PeeringState: do not exclude up from acting\_recovery\_backfill ([pr#31703](#), xie xingguo)
- core: osd/PeeringState: do not start renewing leases until PG is activated ([pr#33129](#), Sage Weil)

- core: osd/PeeringState: fix wrong history of merge target ([pr#29835](#), xie xingguo)
- core: osd/PeeringState: on\_new\_interval on child PG after split ([pr#29780](#), Sage Weil)
- core: osd/PeeringState: recover\_got - add special handler for empty log ([pr#30503](#), xie xingguo)
- core: osd/PeeringState: require SERVER\_OCTOPUS to respond to RenewLease ([pr#33339](#), Neha Ojha)
- core: osd/PeeringState: send pg\_info2 if release >= octopus ([pr#30836](#), Kefu Chai)
- core: osd/PeeringState: transit async\_recovery\_targets back into acting before backfilling ([pr#32202](#), xie xingguo)
- core: osd/PG: Add PG to large omap log message ([pr#30682](#), Brad Hubbard)
- core: osd/PG: adjust pg history on fabricated merge target if necessary ([issue#38623](#), [pr#26822](#), Sage Weil)
- core: osd/PG: clean up fastinfo key when last\_update does not increase ([pr#32615](#), Sage Weil, Kefu Chai)
- core: osd/PG: discover missing objects when an OSD peers and PG is degraded ([pr#27288](#), Jonas Jelten)
- core: osd/PG: do not leak cluster message when theres no con ([pr#32897](#), Sage Weil)
- core: osd/PG: do not queue scrub if PG is not active when unblock ([issue#40451](#), [pr#28660](#), Sage Weil)
- core: osd/PG: do not use approx\_missing\_objects pre-nautilus ([pr#27798](#), Neha Ojha)
- core: osd/PG: fix cleanup of pgmeta-like objects on PG deletion; disallow empty object names ([pr#27929](#), Sage Weil)
- core: osd/PG: fix last\_complete re-calculation on splitting ([issue#26958](#), [pr#27702](#), xie xingguo)
- core: osd/PG: fix \_finish\_recovery vs repair race ([pr#30059](#), xie xingguo)
- core: osd/PG: introduce all\_missing\_unfound helper ([issue#38784](#), [issue#38931](#), [pr#27205](#), xie xingguo)
- core: osd/PG: move down peers out from peer\_purged ([issue#38931](#), [pr#27182](#), xie xingguo)
- core: osd/PG: move } to the proper place ([pr#27204](#), xie xingguo)

- core: osd/PG: remove unused code ([pr#30930](#), Jianpeng Ma)
- core: osd/PG: restart peering for undersized PG on any down stray peer coming back ([pr#33106](#), xie xingguo, Yan Jun)
- core: osd/PG: skip rollforward when !transaction\_applied during append\_log() ([issue#36739](#), [pr#26996](#), Neha Ojha)
- core: osd/PG: the warning seems more serious than what it wanna transmit ([pr#27509](#), Zengran Zhang)
- core: osd/PG: use emplace() to construct new element in-place ([pr#27124](#), Zengran Zhang)
- core: osd/PGLog.h: print olog\_can\_rollback\_to before deciding to rollback ([issue#38894](#), [issue#21174](#), [pr#27105](#), Neha Ojha)
- core: osd/PGLog: persist num\_objects\_missing for replicas when peering is done ([pr#30466](#), xie xingguo)
- core: osd/PGLog: preserve original\_crt to check rollbackability ([issue#36739](#), [pr#27200](#), Neha Ojha)
- core: osd/PGLog: reset log.complete\_to when recover object failed ([pr#30533](#), Tao Ning)
- core: osd/PGStateUtils: initialize NamedState::enter\_time ([pr#33813](#), Jianpeng Ma)
- core: osd/PrimaryLogPG: always use strict priority ordering for kicked recovery ops ([pr#30632](#), xie xingguo)
- core: osd/PrimaryLogPG: Avoid accessing destroyed references in finish\_degrxe2x80xa6 ([pr#29663](#), Tao Ning)
- core: osd/PrimaryLogPG: cancel in-flight manifest ops on interval changing; fix race with scru ([pr#29985](#), xie xingguo)
- core: osd/PrimaryLogPG: do\_op - do not create head object twice ([pr#28785](#), xie xingguo)
- core: osd/PrimaryLogPG: finish\_copyfrom - dirty omap if necessary ([pr#29729](#), xie xingguo)
- core: osd/PrimaryLogPG: fix dirty range of write\_full ([pr#29726](#), xie xingguo)
- core: osd/PrimaryLogPG: fix warning ([pr#30716](#), Sage Weil)
- core: osd/PrimaryLogPG: include op\_returns in dup replies ([pr#30640](#), Sage Weil)
- core: osd/PrimaryLogPG: kill obsolete ondisk\_{read,write}\_lock comments ([pr#29719](#), xie xingguo)

- core: osd/PrimaryLogPG: more constness ([pr#28786](#), Kefu Chai)
- core: osd/PrimaryLogPG: remove unused parent pgls-filter ([pr#29675](#), Radoslaw Zarzynski, Kefu Chai)
- core: osd/PrimaryLogPG: simple debug message ([pr#32444](#), Jianpeng Ma)
- core: osd/PrimaryLogPG: skip obcs that dont exist during backfill scan\_range ([pr#30715](#), Sage Weil)
- core: osd/PrimaryLogPG: update oi.size on write op implicitly truncating object up ([pr#30085](#), xie xingguo)
- core: osd/PrimaryLogPG: use legacy timestamp rendering for hit\_set objects ([pr#33117](#), Sage Weil)
- core: osd/ReplicatedBackend: check against empty data\_included before enabling crc ([pr#29621](#), xie xingguo)
- core: osd/scheduler/OpSchedulerItem: schedule backoffs as client ops ([pr#32382](#), Samuel Just)
- core: osd/SnapMapper: remove pre-octopus snapmapper keys after conversion ([pr#30368](#), Sage Weil)
- core: osd/SnapMirror: no need to record purged\_snaps every epoch ([pr#31866](#), Sage Weil)
- core: OSD: modify n.cookie to op.notify.cookie ([pr#29418](#), yangjun)
- core: osdc/Objecter: always add \0 after strncpy() ([pr#27286](#), Kefu Chai)
- core: osdc/Objecter: Boost.Aasio (I object!) ([pr#16715](#), Adam C. Emerson)
- core: osdc/Objecter: debug pause/unpause transition ([pr#32850](#), Sage Weil)
- core: osdc/Objecter: fix OSDMap leak in handle\_osd\_map ([issue#20491](#), [pr#28242](#), Sage Weil)
- core: osdc/Objecter: only pause if respects\_full() ([pr#33020](#), Sage Weil)
- core: osdc/Objecter: pg-mapping cache ([pr#28487](#), xie xingguo)
- core: osdc/Objecter: \_calc\_target - inline spgid ([pr#28570](#), xie xingguo)
- core: osdc: Fix a missing : for the correct namespace ([pr#29472](#), Willem Jan Withagen)
- core: pybind/ceph\_argparse: improve ceph -h syntax ([pr#30431](#), Sage Weil)
- core: pybind/mgr/mgr\_module: fix standby module logging options ([pr#33639](#), Sage Weil)

- core: pybind/mgr/mgr\_util: fix pretty time delta ([pr#33794](#), Sage Weil)
- core: pybind/mgr/\\*: fix config\_notify handling of default values ([pr#32755](#), Sage Weil)
- core: qa/distros: add rhel/centos 8.1 ([pr#33026](#), Sage Weil)
- core: qa/distros: centos 7.6; update centos and ubuntu latest symlinks ([pr#27349](#), Sage Weil)
- core: qa/standalone/mon/osd-create-pool: fix utf-8 grep LANG ([pr#32711](#), Sage Weil)
- core: qa/standalone/osd/divergent-priors: add reproducer for bug 41816 ([pr#30506](#), Sage Weil)
- core: qa/standalone/osd/osd-bench: debug bluestore ([pr#32961](#), Sage Weil)
- core: qa/standalone/osd/osd-markdown: fix dup command disabling ([issue#38359](#), [pr#27499](#), Sage Weil)
- core: qa/standalone/scrub/osd-scrub-snaps: misc fixes for removed\_snaps change ([issue#40725](#), [pr#29003](#), Sage Weil)
- core: qa/standalone: python -> python3 ([pr#32383](#), Sage Weil)
- core: qa/suites/rados/multimon/tasks/mon\_clock\_with\_skews: disable ntpd etc ([pr#33184](#), Sage Weil)
- core: qa/suites/rados/multimon: fix failures ([issue#40112](#), [pr#28353](#), Sage Weil)
- core: qa/suites/rados/singleton-nomsgr/all/balancer: whitelist PG\_AVAILABILITY ([pr#31747](#), Sage Weil)
- core: qa/suites/rados/singleton/all/ec-lost-unfound: no rbd pool ([pr#30596](#), Sage Weil)
- core: qa/suites/rados/thrash-old-clients: centos -> ubuntu ([pr#32356](#), Sage Weil)
- core: qa/suites/rados/thrash-old-clients: skip TestClsRbd.mirror test ([pr#31745](#), Sage Weil)
- core: qa/suites/rados/thrash: debug monc ([pr#32885](#), Sage Weil)
- core: qa/suites/upgrade/nautilus-x: misc updates ([pr#27138](#), Sage Weil)
- core: qa/suites/upgrade/\\*-x-singleton: enable bluestore debugging settings ([pr#27786](#), Sage Weil)
- core: qa/suites/upgrade: all upgrades to octopus on ubuntu only ([pr#32275](#), Sage Weil)

- core: qa/suits/rados/basic/tasks/rados\_api\_tests: pgs can go degraded ([pr#30627](#), Sage Weil)
- core: qa/tasks/ceph2: teuthology task to bring up a ceph-daemon+ssh cluster ([pr#31502](#), Sage Weil)
- core: qa/tasks/ceph: only re-request scrub on unscrubbed pgs ([pr#32988](#), Sage Weil)
- core: qa/tasks/ceph\_manager: fix thrash\_pg\_upmap\_items when no pools ([pr#29144](#), Sage Weil)
- core: qa/tasks/ceph\_manager: make upmap thrasher behave when no pools/pgs ([pr#29069](#), Sage Weil)
- core: qa/tasks/ceph\_manager: remove race from all\_active\_or\_peered() ([pr#29498](#), Sage Weil)
- core: qa/tasks/ceph\_manager: wait for clean before asserting clean on minsize test ([pr#29109](#), Sage Weil)
- core: qa/workunits/rados/test\_large\_omap\_detection: py3-ify ([pr#32405](#), Sage Weil)
- core: qa: increase mon tell retries when injecting msgr failures ([pr#30872](#), Sage Weil)
- core: qa: more fixes for the removed\_snaps changeset ([issue#40674](#), [pr#28901](#), Sage Weil)
- core: qa: run various tests on ubuntu ([pr#32278](#), Sage Weil)
- core: rados bench: fix the delayed checking of completed ops ([pr#32928](#), Jianshen Liu)
- core: Revert common: default pg\_autoscale\_mode=on for new pools ([pr#30440](#), David Zafman)
- core: Revert crush: remove invalid upmap items ([pr#32017](#), David Zafman)
- core: Revert Merge pull request #16715 from adamemerson/wip-I-Object! ([pr#31790](#), Sage Weil)
- core: Revert test: librados startup/shutdown racer test ([pr#31092](#), Sage Weil)
- core: rgw/rgw\_tools: fix osd pool set json syntax ([pr#27967](#), Sage Weil)
- core: rocksdb: enable rocksdb\_rmrang=true by default ([pr#29323](#), Sage Weil)
- core: rocksdb: Updated to v6.1.2 ([pr#29026](#), Mark Nelson)
- core: sample.ceph.conf: correct the default value of filestore merge threshold ([pr#28653](#), zhang Shaowen)

- core: selinux: Allow ceph to read udev d ([pr#29071](#), Boris Ranto)
- core: src/: Clean up endian handling ([pr#30409](#), Ulrich Weigand)
- core: src/dmclock: bring in fixes for indirect\_intrusive\_heap ([pr#32380](#), Samuel Just)
- core: src/osd: add tier-flush op ([pr#28778](#), Myoungwon Oh)
- core: test: add librados-based startup/shutdown racer test ([pr#30552](#), Jeff Layton)
- core: tools/rados: call pool\_lookup() after rados is connected ([pr#30413](#), Vikhyat Umrao)
- core: tools/rados: prevent put operation from recreating object when -offset=0 ([pr#31230](#), Adam Kupczyk)
- core: tools/rados: Unmask -o to restore original behaviour ([pr#31310](#), Brad Hubbard)
- core: Wip lazy omap test ([pr#28070](#), Brad Hubbard)
- crimson/osd: serve read requests ([pr#26697](#), Kefu Chai)
- Crimson build fixes ([pr#33345](#), Samuel Just)
- crimson, common: Add ephemeral ObjectContext state to crimson ([pr#31202](#), Samuel Just)
- crimson, auth: fix FTBFS of crimson-osd and fix v1/v2 auth ([pr#27809](#), Kefu Chai, Yingxin Cheng)
- crimson, osd: performance fixes ([pr#28071](#), Kefu Chai, Radoslaw Zarzynski)
- crimson/common/errorator.h: add handle\_error() method ([pr#31856](#), Radoslaw Zarzynski)
- crimson/common/errorator.h: simplify the compound safe\_then() variant ([pr#31918](#), Radoslaw Zarzynski)
- crimson/common: more friendly to seastar::do\_with() ([pr#33199](#), Kefu Chai)
- crimson/common: remove unused file .#log.cc ([pr#28828](#), Changcheng Liu)
- crimson/mon: fix the v1 auth ([pr#28041](#), Kefu Chai)
- crimson/mon: use shared\_future for waiting MauthReply ([pr#30366](#), chunmei Liu)
- crimson/net: bug fixes from v2 failover tests ([pr#29882](#), Yingxin Cheng)
- crimson/net: clean-up and fixes of messenger ([pr#29057](#), Yingxin Cheng)

- crimson/net: extract do\_write\_dispatch\_sweep() ([pr#27428](#), Yingxin Cheng)
- crimson/net: implement preemptive shutdown/close ([pr#28682](#), Yingxin Cheng)
- crimson/net: improve batching in the write path ([pr#27788](#), Yingxin Cheng)
- crimson/net: lossless policy for v2 protocol ([pr#29378](#), Yingxin Cheng)
- crimson/net: lossy connection for ProtocolV2 ([pr#26710](#), Yingxin Cheng)
- crimson/net: misc fixes in v1 read path ([pr#27837](#), Yingxin Cheng)
- crimson/net: prefer <fmt/chrono.h> over <fmt/time.h> ([pr#27831](#), Kefu Chai)
- crimson/net: prevent reusing the sent messages ([pr#28890](#), Yingxin Cheng)
- crimson/net: print tx/rx messages using logger().info() ([pr#28798](#), Kefu Chai)
- crimson/net: remove redundant std::move() ([pr#28317](#), Kefu Chai)
- crimson/net: v2 racing tests, stall tests and bug fixes ([pr#30313](#), Yingxin Cheng)
- crimson/os: do not fail if fsid file exists when mkfs ([pr#27006](#), chunmei Liu, Kefu Chai)
- crimson/os: init PG with pg coll not meta coll ([pr#33084](#), Kefu Chai)
- crimson/os: Object::read() returns bufferlist instead of never used errcode ([pr#30380](#), Radoslaw Zarzynski)
- crimson/osd/osd\_operation.h: clean up duplicative check ([pr#31859](#), Radoslaw Zarzynski)
- crimson/osd/pg: start\_operation for read\_state, schedule\_event\_on\_commit ([pr#28771](#), Samuel Just)
- crimson/osd/pg\_meta: use initializer list for passing set<> ([pr#28461](#), Kefu Chai)
- crimson/osd: abort on unsupported objectstore type ([pr#28790](#), Kefu Chai)
- crimson/osd: add -help-seastar command line option ([pr#28794](#), Kefu Chai)
- crimson/osd: add minimal state machine for PG peering ([pr#27071](#), Kefu Chai)
- crimson/osd: add pgl support ([pr#30433](#), Kefu Chai)
- crimson/osd: cache object\_info and snapset in PGBackend ([pr#27310](#), Kefu Chai)
- crimson/osd: call at\_exit() before stopping the engine ([pr#27177](#), Kefu Chai)
- crimson/osd: call engine().exit(0) after mkfs ([pr#27061](#), Kefu Chai)
- crimson/osd: capture watcher when calling its member function ([pr#33425](#), Kefu Chai)

Chai)

- crimson/osd: cleanups ([pr#30736](#), Kefu Chai)
- crimson/osd: consolidate the code to initialize msgrs ([pr#27426](#), Kefu Chai)
- crimson/osd: create msgrs in main.cc ([pr#27066](#), Kefu Chai)
- crimson/osd: crimson/osd: do not load fullmap.0 ([pr#27004](#), chunmei Liu, Kefu Chai)
- crimson/osd: differentiate write from writefull ([pr#28959](#), Kefu Chai)
- crimson/osd: do not add whoami as hb peer and cleanups ([pr#27307](#), Kefu Chai)
- crimson/osd: extend OpsExecuter to carry about op effects ([pr#30310](#), Radoslaw Zarzynski)
- crimson/osd: fix the build broken by df771861 ([pr#28053](#), chunmei Liu)
- crimson/osd: fix the Clang build in create\_watch\_info() ([pr#33350](#), Radoslaw Zarzynski)
- crimson/osd: implement replicated write ([pr#29076](#), Kefu Chai)
- crimson/osd: init PG with more info ([pr#27064](#), Kefu Chai)
- crimson/osd: lower debug level on i/o path ([pr#27338](#), Kefu Chai)
- crimson/osd: misc fixes and cleanup ([pr#33528](#), Yingxin Cheng)
- crimson/osd: misc fixes for OSD reboot-ability ([pr#33595](#), Yingxin Cheng)
- crimson/osd: partition args the right way ([pr#27211](#), Kefu Chai)
- crimson/osd: pass unknown args to ConfigProxy::parse\_args() ([pr#27062](#), Kefu Chai)
- crimson/osd: remove unneeded captures - pg.cc ([pr#33349](#), Ronen Friedman)
- crimson/osd: report pg\_stats to mgr ([pr#27065](#), Kefu Chai)
- crimson/osd: should handle pg\_lease messages ([pr#30834](#), Kefu Chai)
- crimson/osd: shutdown services in the right order ([pr#27987](#), Kefu Chai)
- crimson/osd: some cleanups ([pr#28402](#), Kefu Chai)
- crimson/osd: support write pid\_file when osd start ([pr#27413](#), chunmei Liu)
- crimson/osd: update peering\_state in PG::on\_activate\_complete() ([pr#28747](#), Kefu Chai)
- crimson/osd: use single-pg peering ops ([pr#30372](#), Kefu Chai)

- crimson/thread: generalize Task so it works w/ func returns void ([pr#32742](#), Kefu Chai)
- crimson/{net,mon,osd}: misc logging changes ([pr#27099](#), Kefu Chai)
- crimson/{osd,heartbeat}: allow heartbeat to have access to authorizer ([pr#27059](#), Kefu Chai)
- crimson/{osd,mon}: lower log level when sending a replicated op ([pr#30957](#), Kefu Chai)
- crimson: add editor properties header ([pr#33408](#), Kefu Chai)
- crimson: add FuturizedStore to encapsulate CyanStore ([pr#28358](#), chunmei Liu)
- crimson: add missing include in common/errorator.h ([pr#32490](#), Radoslaw Zarzynski)
- crimson: add support for basic write path ([pr#27873](#), Radoslaw Zarzynski)
- crimson: add support for watch / notify, part 1 ([pr#32679](#), Radoslaw Zarzynski)
- crimson: bring ceph::errorator with its first appliances ([pr#30387](#), Radoslaw Zarzynski)
- crimson: CLANG-related fixes to errorator.h ([pr#32488](#), Ronen Friedman, Radoslaw Zarzynski)
- crimson: clean up and refactor asok ([pr#33357](#), Kefu Chai)
- crimson: enable cephx for v2 msgr ([pr#27514](#), Kefu Chai)
- crimson: fix build with GCC-10 ([pr#33233](#), Kefu Chai)
- crimson: fix crimson pg coll usage error ([pr#33076](#), Chunmei Liu)
- crimson: fix lambda captures of non-variables ([pr#32494](#), Ronen Friedman)
- crimson: futurized CyanStores member functions and Collection ([pr#29470](#), Kefu Chai, chunmei Liu)
- crimson: handle MOSDPGQuery2 properly ([pr#30399](#), Kefu Chai)
- crimson: make seastar::do\_with() a friend of errorated futures ([pr#32175](#), Radoslaw Zarzynski)
- crimson: move dummy impl of AuthServer to DummyAuth ([pr#27452](#), Kefu Chai)
- crimson: move os/cyan\\_\\* down to os/cyanstore/\\* ([pr#31874](#), Kefu Chai)
- crimson: pass Connection\\* to Dispatch::ms\_dispatch() ([pr#27690](#), Yingxin Cheng, Kefu Chai)
- crimson: pickup change to fix -cpuset support and cleanups ([pr#33250](#), Kefu Chai)

- crimson: remove some attributes from lambda ([pr#32604](#), Ronen Friedman)
- crimson: run in foreground if possible, silence warnings ([pr#30474](#), Samuel Just, Kefu Chai)
- crimson: s/ceph/crimson/ in namespace names ([pr#31069](#), Kefu Chai)
- crimson: serve basic RBD traffic coming from fio ([pr#30339](#), Radoslaw Zarzynski)
- crimson: solve the problem that crimson-osds created pgs stuck in unknown state ([pr#33780](#), Xuehan Xu)
- crimson: stop osd before stopping messengers ([pr#31904](#), Kefu Chai)
- crimson: support pgnls and delete op ([pr#28079](#), Kefu Chai)
- crimson: update osd when peer gets authenticated ([pr#27416](#), Kefu Chai)
- crimson: use given osd\_fsid when mkfs ([pr#28800](#), Kefu Chai)
- crimson:: add alien blue store ([pr#31041](#), Samuel Just, Chunmei Liu, Kefu Chai)
- crush: add root\_bucket to identify underfull buckets ([issue#38826](#), [pr#27068](#), huangjun)
- crush: remove invalid upmap items ([pr#31131](#), huangjun)
- crush: remove invalid upmap items ([pr#32099](#), huangjun)
- crush: various fixes for weight-sets, the osd\_crush\_update\_weight\_set option, and tests ([pr#26955](#), Sage Weil)
- dashboard/services: fix lint error ([pr#30289](#), Willem Jan Withagen)
- deb,rpm: switch to python 3 ([pr#32252](#), Sage Weil, Alfredo Deza)
- debian: add python3-jsonpatch as dependency ([pr#33298](#), Sebastian Wagner)
- denc: allow DencDumper to dump OOB buffer ([pr#27704](#), Kefu Chai)
- doc/bootstrap: fixed default -keyring target ([pr#32643](#), Yaarit Hatuka)
- doc/foundation: fix amihan ([pr#32999](#), Sage Weil)
- doc: .organizationmap: Wido 42on -> 42on ([pr#32260](#), Sage Weil)
- doc: add a deduplication document ([pr#28462](#), Myoungwon Oh)
- doc: add a doc for vstart\_runner.py ([pr#29907](#), Rishabh Dave)
- doc: add a new document on distributed cephfs metadata cache ([pr#30265](#), Jeff Layton)

- doc: Add a new document on Dynamic Metadata Management in CephFS ([pr#30348](#), Sidharth Anupkrishnan)
- doc: Add a RGW swift auth note ([pr#31309](#), Matthew Oliver)
- doc: add ceph fs volumes and subvolumes documentation ([pr#30381](#), Ramana Raja)
- doc: add CephFS Octopus release notes ([pr#33450](#), Patrick Donnelly)
- doc: add changelog for nautilus ([pr#27048](#), Abhishek Lekshmanan)
- doc: add chrony to preflight checklist for Ubuntu 18.04 ([pr#31948](#), Zac Dover)
- doc: add config help/get/set section for runtime client configuration ([issue#41688](#), [pr#32117](#), Venky Shankar)
- doc: Add Dashboard Octopus release notes ([pr#33555](#), Lenz Grimmer)
- doc: add description for fuse\_disable\_pagecache ([pr#31902](#), Yan, Zheng)
- doc: add doc for blacklisting older CephFS clients ([issue#39130](#), [pr#27412](#), Patrick Donnelly)
- doc: add doc for cephfs lazyio ([issue#38729](#), [pr#26976](#), Yan, Zheng)
- doc: add guide for running tests with teuthology ([pr#32114](#), Rishabh Dave)
- doc: add mds map to list of ceph monitor assets ([pr#32631](#), Zac Dover)
- doc: add missed word than in doc/man/8/rbd.rst ([pr#31022](#), Drunkard Zhang)
- doc: Add missing mgr cap for the bootstrap keyring ([pr#27201](#), Bryan Stillwell)
- doc: add missing virtualenv for build-doc ([pr#31896](#), Rodrigo Severo)
- doc: Add note to execute cephfs-shell ([pr#27369](#), Varsha Rao)
- doc: add package for Golang ([issue#38730](#), [pr#26937](#), Irek Fasikhov)
- doc: add Python 2 to Ubuntu 18.04 installations ([pr#31947](#), Zac Dover)
- doc: add release notes for 13.2.5 mimic ([pr#26913](#), Abhishek Lekshmanan)
- doc: add release notes for v13.2.6 mimic ([pr#28385](#), Abhishek Lekshmanan)
- doc: Add sphinx\_autodoc\_typehints extension ([pr#33577](#), Sebastian Wagner)
- doc: Add stat command usage in cephfs-shell ([pr#28236](#), Varsha Rao)
- doc: Add usage for shortcuts command in cephfs-shell ([pr#27373](#), Varsha Rao)
- doc: Add warning that the root directory cannot be fragmented ([pr#28354](#), Nathan Fish)

- doc: Added a link to Ceph Community Calendar ([pr#31475](#), Zac Dover)
- doc: added a remark to always use powers of two for pg\_num ([pr#31541](#), Thomas Schneider)
- doc: added an is where it was needed ([pr#32374](#), Zac Dover)
- doc: Added dashboard features, improved wording ([pr#27997](#), Lenz Grimmer)
- doc: added section on creating RESTful API user ([pr#26016](#), James McClune)
- doc: Added the crisp getting started guide to index.rst ([pr#32531](#), Zac Dover)
- doc: Adding US-Mid-West Mirror to docs ([pr#25099](#), Mike Perez)
- doc: Adds cmake build options for optionally skipping few components ([pr#31066](#), Deepika Upadhyay)
- doc: adjust for mon\_status changes in octopus ([pr#33703](#), Nathan Cutler)
- doc: admin/doc/\_ext/ceph\_releases.py: use yaml.safe\_load() ([pr#28463](#), Kefu Chai)
- doc: admin/build-doc: always install python3-\* for build deps ([pr#32481](#), Kefu Chai)
- doc: admin/build-doc: do not use system site-packages ([pr#32285](#), Sage Weil)
- doc: admin/build-doc: Fix doxygen typo ([pr#32572](#), Varsha Rao)
- doc: admin/build-doc: use python3 ([pr#29528](#), Kefu Chai)
- doc: admin/doc-requirements.txt: bump up Sphinx and breathe ([pr#32301](#), Kefu Chai)
- doc: admin/serve-doc: Switch to python3 only ([pr#33596](#), Brad Hubbard)
- doc: always load resources via HTTPS ([pr#29544](#), Tiago Melo)
- doc: ceph-monstore-tool: correct the key for storing mgr\_command\_descs ([pr#33172](#), Kefu Chai)
- doc: cephfs: add section on fsync error reporting to posix.rst ([issue#24641](#), [pr#28300](#), Jeff Layton)
- doc: change case from apis to APIs ([pr#33664](#), Deepika Upadhyay)
- doc: clarify difference between fs and kcephfs suite ([pr#32144](#), Rishabh Dave)
- doc: clarify priority use ([pr#32191](#), Yuri Weinstein)
- doc: clarify support for rbd fancy striping ([pr#32176](#), Ilya Dryomov)
- doc: cleanup CephFS Landing Page ([pr#30542](#), Milind Changire)

- doc: coding-style: update a link and fix typos ([pr#33128](#), Ponnuel Palaniyappan)
- doc: common/admin\_socket: Add doxygen for call and call\_async ([pr#32547](#), Adam Kupczyk)
- doc: common/hobject: Error invocation of formula in documentation ([pr#28366](#), Albert)
- doc: config-ref: add a note on current scheduler settings ([pr#27243](#), Abhishek Lekshmanan)
- doc: correct example to use vstart to run up cluster ([pr#26816](#), Changcheng Liu)
- doc: cover more cache modes in rados/operations/cache-tiering.rst ([issue#14153](#), [pr#17614](#), Nathan Cutler)
- doc: default values for mon\_health\_to\_clog\\_\* were flipped ([pr#29867](#), James McClune)
- doc: describe metadata\_heap cleanup ([issue#18174](#), [pr#26915](#), Dan van der Ster)
- doc: Describe recovery and backfill prioritizations ([issue#39011](#), [pr#27941](#), David Zafman)
- doc: doc : fixed capitalization ([pr#27379](#), Servesha Dudhgaonkar)
- doc: doc, qa: remove invalid option mon\_pg\_warn\_max\_per\_osd ([pr#30787](#), zhang daolong)
- doc: doc,admin: fix the builtin search ([pr#33592](#), Kefu Chai)
- doc: doc/architecture.rst: fix a typo in EC section ([pr#33241](#), Nag Pavan Chilakam)
- doc: doc/bootstrap.rst: fix githus url ([pr#31086](#), Alexandre Bruyelles)
- doc: doc/bootstrap: add mds and rgw steps to bootstrap ([pr#33088](#), Sage Weil)
- doc: doc/ceph-fuse: describe -n option ([pr#30911](#), Rishabh Dave)
- doc: doc/ceph-fuse: mention -k option in ceph-fuse man page ([pr#30561](#), Rishabh Dave)
- doc: doc/ceph-kvstore-tool: add description for stats command ([pr#29990](#), Josh Durgin, Adam Kupczyk)
- doc: doc/ceph-volume: initial docs for zfs/inventory and zfs/api ([pr#31252](#), Willem Jan Withagen)
- doc: doc/cephadm/administration: clarify log gathering ([pr#33627](#), Nathan Cutler)
- doc: doc/cephadm: adjust syntax for config set ([pr#33600](#), Joshua Schmid)

- doc: doc/cephadm: big cleanup of cephadm docs ([pr#33981](#), Sage Weil)
- doc: doc/cephadm: Troubleshooting ([pr#33460](#), Sebastian Wagner)
- doc: doc/cephfs/client-auth: description and example are inconsistent ([pr#32762](#), Ilya Dryomov)
- doc: doc/cephfs/disaster-recovery-experts: Add link for scrub and note for scrub\_path ([pr#32124](#), Varsha Rao)
- doc: doc/cephfs: add doc for cephfs io path ([pr#30369](#), Yan, Zheng)
- doc: doc/cephfs: correct a description mistake about mds states ([issue#41893](#), [pr#30427](#), Xiao Guodong)
- doc: doc/cephfs: improve add/remove MDS section ([issue#39620](#), [pr#28700](#), Patrick Donnelly)
- doc: doc/cephfs: migrate best practices recommendations to relevant docs ([pr#32522](#), Rishabh Dave)
- doc: doc/cleanup: drop repo-access.rst ([pr#32276](#), Nathan Cutler)
- doc: doc/corpus: update to adapt the change from autotools to cmake ([pr#27552](#), Kefu Chai)
- doc: doc/dev/corpus.rst: correct instructions ([pr#27741](#), Kefu Chai)
- doc: doc/dev/corpus.rst: minor tweaks ([pr#28877](#), Kefu Chai)
- doc: doc/dev/crimson.rst: document CBT testing ([pr#30290](#), Kefu Chai)
- doc: doc/dev/crimson: transpose options of compare.py ([pr#30453](#), Kefu Chai)
- doc: doc/dev/developer\_guide/index.rst: add youtube reference for Getting Started ([pr#29712](#), Neha Ojha)
- doc: doc/dev/developer\_guide/index.rst: add youtube references ([pr#29033](#), Neha Ojha)
- doc: doc/dev/developer\_guide: fix heading level ([pr#30428](#), Nathan Cutler)
- doc: doc/dev/developer\_guide: remove web address ([pr#29183](#), gabriellasroman)
- doc: doc/dev/kubernetes: Update ([pr#28081](#), Sebastian Wagner)
- doc: doc/dev/osd\_internals/async\_recovery: update cost calculation ([pr#28036](#), Neha Ojha)
- doc: doc/dev: add crimson.rst ([pr#28674](#), Kefu Chai)
- doc: doc/dev: add teuthology priority recommendations ([pr#30308](#), Patrick

Donnelly)

- doc: doc/developer: fix dev mailing list address ([pr#32442](#), Willem Jan Withagen)
- doc: doc/drivegroups: add docs for DriveGroups with excessive examples ([pr#33044](#), Joshua Schmid)
- doc: doc/foundation: add ceph foundation info here ([pr#31955](#), Sage Weil)
- doc: doc/foundation: add cloudbase and vexxhost ([pr#32013](#), Sage Weil)
- doc: doc/foundation: add Samsung Electronics ([pr#33518](#), Sage Weil)
- doc: doc/governance: add cbodey ([pr#27708](#), Sage Weil)
- doc: doc/index: remove quick start from front page for now ([pr#33207](#), Sage Weil)
- doc: doc/install/containers: add summary of containers and branches ([pr#31465](#), Sage Weil)
- doc: doc/install/containers: note vX.Y.Z[-YYYYMMDD] tags ([pr#31975](#), Sage Weil)
- doc: doc/install/manual-deployment: Change owner to ceph for the keyring file ([pr#31452](#), Jeffrey Chu)
- doc: doc/install/upgrading-ceph: systemctl in Ubuntu instructions ([pr#32595](#), Rodrigo Severo)
- doc: doc/install: rethink install doc installation methods order ([pr#33890](#), Zac Dover, Sebastian Wagner)
- doc: doc/man/ceph: document ceph config ([pr#30645](#), Kefu Chai)
- doc: doc/man: improve bluefs-bdev-expand option ([pr#32590](#), Kefu Chai)
- doc: doc/mgr/ansible.rst: fix typo ([pr#28827](#), Lan Liu)
- doc: doc/mgr/cephadm: document adoption process ([pr#33459](#), Sage Weil)
- doc: doc/mgr/orchestrator.rst: updated current implementation status ([pr#33410](#), Kai Wagner)
- doc: doc/mgr/orchestrator: Add Cephfs ([pr#33574](#), Sebastian Wagner)
- doc: doc/mgr/orchestrator\_cli: Rook orch supports mon update ([issue#39137](#), [pr#27431](#), Sebastian Wagner)
- doc: doc/mgr/telemetry: added device channel details ([pr#33113](#), Yaarit Hatuka)
- doc: doc/mgr/telemetry: update default interval ([pr#31008](#), Tim Serong)
- doc: doc/mgr: Enhance placement specs ([pr#33924](#), Sebastian Wagner)

- doc: doc/orchestrator: Fix broken bullet points ([issue#39094](#), [pr#27121](#), Sebastian Wagner)
- doc: doc/orchestrator: Fix various issues in Orchestrator CLI documentation ([pr#31353](#), Volker Theile)
- doc: doc/orchestrator: Sync status with reality ([pr#30281](#), Sebastian Wagner)
- doc: doc/orchestrator: update rgw creation ([pr#33540](#), Yehuda Sadeh)
- doc: doc/rados/api/python: Add documentation for mon\_command ([pr#26934](#), Sebastian Wagner)
- doc: doc/rados/configuration/osd-config-ref.rst: document osd\_delete\_sleep ([pr#28775](#), Neha Ojha)
- doc: doc/rados/configuration: fix typo in mon-lookup-dns ([pr#27362](#), Vanush Misha Paturyan)
- doc: doc/rados/configuration: fix typos in osd-config-ref.rst ([pr#28805](#), Lan Liu)
- doc: doc/rados/configuration: update to be in sync with ConfUtils changes ([pr#28753](#), Kefu Chai)
- doc: doc/rados/deployment/ceph-deploy-mon: fix typo ([pr#31164](#), Kefu Chai)
- doc: doc/rados/operations/crush-map-edits: recompile and set instructions ([pr#32451](#), Rodrigo Severo)
- doc: doc/rados/operations/devices: document device failure prediction ([pr#27472](#), Sage Weil)
- doc: doc/rados/operations/erasure-code.rst: allow recovery below min\_size ([pr#28750](#), Greg Farnum, Neha Ojha)
- doc: doc/rados/operations: add safe-to-destroy check to OSD replacement workflow ([pr#28491](#), Sage Weil)
- doc: doc/rados/operations: crush\_rule is a name ([pr#29367](#), Kefu Chai)
- doc: doc/rados/operations: document BLUEFS\_SPILLOVER ([pr#27316](#), Sage Weil)
- doc: doc/rados/operations: min\_size is applicable to EC ([pr#33543](#), Brad Hubbard)
- doc: doc/rados/operations: OSD\_OUT\_OF\_ORDER\_FULL fullness order is wrong ([pr#31588](#), Tsung-Ju Lii)
- doc: doc/rados: Better block.db size recommendations for bluestore ([pr#32226](#), Neha Ojha)
- doc: doc/rados: Correcting some typos in the clay code documentation ([pr#29889](#), Myna)

- doc: doc/rados: update osd\_min\_pg\_log\_entries and add osd\_max\_pg\_log\_entries ([pr#32790](#), Neha Ojha)
- doc: doc/radosgw/admin:fix how to modify subuser info ([pr#29839](#), Feng Hualong)
- doc: doc/radosgw/compression.rst: fix typo ([pr#28749](#), hydro-)
- doc: doc/radosgw/config-ref: paragraph to explain the gc settings ([pr#32367](#), Kai Wagner)
- doc: doc/radosgw/multisite-sync-policy.rst: fix typo ([pr#33230](#), Liu Lan)
- doc: doc/radosgw: fix typos ([pr#30642](#), Liu Lan)
- doc: doc/radosgw: update documentation examples with the current S3 PHP client ([pr#25985](#), Laurent VOULEMIER)
- doc: doc/rbd/rbd-cloudstack: update disk offering URL to new docs ([pr#27713](#), Kefu Chai)
- doc: doc/rbd: document the new snapshot-based mirroring feature ([pr#33561](#), Jason Dillaman)
- doc: doc/rbd: fix small typos ([pr#33689](#), songweibin)
- doc: doc/rbd: initial kubernetes / ceph-csi integration documentation ([pr#29429](#), Jason Dillaman)
- doc: doc/rbd: re-organize top-level and add live-migration docs ([issue#40486](#), [pr#29135](#), Jason Dillaman)
- doc: doc/rbd: refine rbd/libvirt usage ([pr#32273](#), Changcheng Liu)
- doc: doc/rbd: s/guess/xml/ for codeblock lexer ([pr#30953](#), Kefu Chai)
- doc: doc/rbd: simplify libvirt usage ([pr#32142](#), Changcheng Liu)
- doc: doc/rbd: update krbd version support for RBD features ([issue#40802](#), [pr#29083](#), Jason Dillaman)
- doc: doc/release/nautilus: 14.2.2 changes redone ([pr#29145](#), Sage Weil)
- doc: doc/release/octopus: note about upgrade times ([pr#33401](#), Sage Weil)
- doc: doc/releases/nautilus,PendingReleaseNotes: consolidate telemetry note ([pr#32160](#), Sage Weil)
- doc: doc/releases/nautilus.rst: fix command to check min\_compatible\_client ([pr#28526](#), Osama Elswah)
- doc: doc/releases/nautilus.rst: remove a redundant \\* ([pr#32577](#), Servesha Dudhgaonkar)

- doc: doc/releases/nautilus: Correct a systemctl command in an upgrade guide ([pr#27773](#), Teeranai Kormongkolkul)
- doc: doc/releases/nautilus: final notes for v14.2.0 ([pr#27019](#), Sage Weil)
- doc: doc/releases/nautilus: fix config update step ([pr#27495](#), Sage Weil)
- doc: doc/releases/nautilus: fix release notes (crash->device) ([pr#32148](#), Sage Weil)
- doc: doc/releases/octopus.rst: add note about ec recovery below min\_size ([pr#34092](#), Neha Ojha)
- doc: doc/releases/octopus.rst: format tweaks ([pr#33971](#), Kefu Chai)
- doc: doc/releases/octopus.rst: formatting tweaks ([pr#33987](#), Kefu Chai)
- doc: doc/releases/octopus: add additional RBD improvements ([pr#34032](#), Jason Dillaman)
- doc: doc/releases/schedule.rst: add 14.2.3, 14.2.4, 15.0.0 and drop dumpling ([pr#30430](#), Nathan Cutler)
- doc: doc/releases: access main releases page from top-level TOC ([pr#30598](#), Nathan Cutler)
- doc: doc/releases: add 14.2.8 to release timeline ([pr#33721](#), Nathan Cutler)
- doc: doc/releases: add mimic v13.2.7 to releases timeline ([pr#31872](#), Nathan Cutler)
- doc: doc/releases: add release notes for mimic v13.2.7 ([pr#31777](#), Nathan Cutler)
- doc: doc/releases: add release notes for mimic v13.2.8 ([pr#32040](#), Nathan Cutler)
- doc: doc/releases: add release notes for nautilus v14.2.5 ([pr#31970](#), Nathan Cutler)
- doc: doc/releases: Ceph Nautilus v14.2.4 Release Notes ([pr#30429](#), Nathan Cutler)
- doc: doc/releases: octopus draft notes ([pr#33043](#), Sage Weil)
- doc: doc/releases: Octopus is not stable yet ([pr#33729](#), Nathan Cutler)
- doc: doc/releases: update for 12 month cycle ([pr#28864](#), Sage Weil)
- doc: doc/rgw: add design doc for multisite resharding ([pr#33539](#), Casey Bodley)
- doc: doc/rgw: document CreateBucketConfiguration for s3 PUT Bucket api ([issue#39597](#), [pr#27977](#), Casey Bodley)
- doc: doc/rgw: document use of realm pull instead of period pull ([issue#39655](#),

[pr#28052](#), Casey Bodley)

- doc: doc/rgw: fix broken link to boto s3 extensions document ([pr#32740](#), Casey Bodley)
- doc: doc/rgw: update civetweb rgw\_frontends config example ([pr#27054](#), Casey Bodley)
- doc: doc/start/documenting-ceph.rst: make better doc recommendations ([pr#30273](#), Neha Ojha)
- doc: doc/start/hardware-recommendations.rst: minor tweaks ([pr#30837](#), Amrita Sakthivel)
- doc: doc/\_templates/page.html: redirect to etherpad ([pr#32197](#), Neha Ojha)
- doc: Doc: Add Nautilus 14.2.2 to schedule and releases ([issue#40988](#), [pr#29362](#), JuanJose Galvez)
- doc: Doc: update release schedule ([pr#28466](#), Torben Hxc3xb8rup)
- doc: docs: fix rgw\_ldap\_dnattr username token ([pr#27964](#), Thomas Kriechbaumer)
- doc: docs: improve rgw ldap auth options ([pr#28157](#), Thomas Kriechbaumer)
- doc: docs: rgw: fix bucket operation spelling: ListBucketMultipartUploads ([pr#28885](#), Thomas Kriechbaumer)
- doc: docs: Update au.ceph.com maintainers, update README.md ([pr#32814](#), Matthew Taylor)
- doc: Document Export Process during Subtree Migrations ([pr#30751](#), Sidharth Anupkrishnan)
- doc: document mds journal event types ([issue#42190](#), [pr#30749](#), Venky Shankar)
- doc: document mds journaling ([issue#41783](#), [pr#30396](#), Venky Shankar)
- doc: document mode param for rbd mirror image enable command ([pr#32735](#), Mykola Golub)
- doc: document rank option for journal reset ([pr#31201](#), Patrick Donnelly)
- doc: document the new -addv argument ([issue#40568](#), [pr#28819](#), Luca Castoro)
- doc: Documentation: Add missing ceph-volume lvm batch argument to ceph-volume.rst ([pr#29081](#), Andreas Krebs)
- doc: Documentation: Centos ceph-deploys python dependencies ([pr#32591](#), C1xc3xa9ment Hampaxc3xaf)
- doc: documentation: Updated Dashboard Features, improved flow ([pr#33919](#), Lenz

Grimmer)

- doc: drop and update troubleshooting ([pr#28900](#), Jos Collin)
- doc: emphasize the importance of require-osd-release nautilus ([pr#32587](#), Zac Dover)
- doc: fix a typo in a command ([pr#32230](#), taeuk\_kim)
- doc: Fix a typo in balancer documentation ([pr#30210](#), Francois Deppierraz)
- doc: fix boot transition in mds state diagram ([pr#27685](#), Patrick Donnelly)
- doc: fix errors in search page and use relative address for releases.json ([pr#33423](#), Kefu Chai)
- doc: Fix for new ceph-devel mailing list ([pr#29492](#), David Zafman)
- doc: Fix FUSE expansion ([pr#30473](#), Sidharth Anupkrishnan)
- doc: fix Getting Started with CephFS ([pr#32457](#), Jos Collin)
- doc: fix links in developer\_guide ([pr#32728](#), Rishabh Dave)
- doc: fix LRC documentation ([pr#27106](#), Danny Al-Gaaf)
- doc: fix parameter to set pg autoscale mode ([pr#27422](#), Changcheng Liu)
- doc: Fix rbd namespace documentation ([pr#29445](#), Ricardo Marques)
- doc: Fix the pg states and auto repair config options ([issue#38896](#), [pr#27143](#), David Zafman)
- doc: fix typo ([pr#28888](#), Jos Collin)
- doc: fix typo in doc/radosgw/layout.rst ([pr#29932](#), ypdai)
- doc: fix typo to auto scale pg number ([pr#31065](#), Changcheng Liu)
- doc: fix typos ([pr#30583](#), Michael Prokop)
- doc: fix urls ([pr#29300](#), Jos Collin)
- doc: fixed -read-only argument value in multisite doc ([pr#28655](#), Chenjiong Deng)
- doc: fixed broken link in Swift Settings section ([pr#28774](#), James McClune)
- doc: fixed broken links in nautilus release page ([pr#28074](#), James McClune)
- doc: fixed broken reference link for Graphviz ([pr#32021](#), James McClune)
- doc: fixed caps ([pr#27397](#), Servesha Dudhgaonkar)
- doc: fixed telemetry module reference link ([pr#27624](#), James McClune)

- doc: fixed typo in leadership names ([pr#27396](#), Servesha Dudhgaonkar)
- doc: Fixes OSD node labels which based on the osd\_devices name ([pr#23312](#), Siyu Sun)
- doc: Fixes typo for ceph dashboard command ([pr#30292](#), Fabian Bonk)
- doc: hide page contents for Ceph Internals ([pr#31046](#), Milind Changire)
- doc: improve ceph-backport.sh comment block ([pr#28042](#), Nathan Cutler)
- doc: improve developer guide doc ([pr#30435](#), Rishabh Dave)
- doc: improve in mount.ceph man page ([pr#31024](#), Rishabh Dave)
- doc: Improved the dashboard proxy config section ([pr#27581](#), Lenz Grimmer)
- doc: indicate imperative mood for commit titles ([pr#29509](#), Patrick Donnelly)
- doc: Make ceph-dashboard require grafana dashboards ([pr#28997](#), Boris Ranto)
- doc: mds-config-ref: update mds\_log\_max\_segments value ([pr#29412](#), Konstantin Shalygin)
- doc: mention -namespace option in rados manpage ([pr#31871](#), Nathan Cutler)
- doc: mgr/dashboard: Add frontend code documentation ([issue#36243](#), [pr#27433](#), Ernesto Puerta)
- doc: mgr/dashboard: Document UiApiController with ApiController usage ([pr#29819](#), Stephan Mxc3xbcller)
- doc: mgr/dashboard: Extend Writing End-to-End Tests section (describe vs it) ([pr#29707](#), Adam King, Rafael Quintero)
- doc: mgr/dashboard: fix hacking.rst ([pr#27222](#), Ernesto Puerta)
- doc: mgr/dashboard: Fix link format to HACKING.rst ([pr#28897](#), Ernesto Puerta)
- doc: mgr/dashboard: fix typos in HACKING.rst ([pr#30847](#), Ernesto Puerta)
- doc: mgr/orchestrator: Add error handling to interface ([pr#26404](#), Sebastian Wagner)
- doc: mgr/orchestrator: Fix disabling the orchestrator ([issue#40779](#), [pr#29042](#), Sebastian Wagner)
- doc: mgr/orchestrator\_cli: Update doc link in README ([pr#31731](#), Varsha Rao)
- doc: mgr/ssh: HACKING.rst: Add Understanding AsyncCompletion ([pr#31967](#), Sebastian Wagner)
- doc: mgr/ssh: update ssh-orch bootstrap guide (Vagrantfile & docs) ([pr#31457](#),

Joshua Schmid)

- doc: mgr/telemetry: force --license when sending while opted-out ([pr#33747](#), Yaarit Hatuka)
- doc: minor fix in mount.ceph ([pr#32748](#), Rishabh Dave)
- doc: Miscellaneous spelling fixes ([pr#27202](#), Bryan Stillwell)
- doc: Modify nature theme ([pr#32312](#), Brad Hubbard)
- doc: mon/OSDMonitor: Fix pool set target\_size\_bytes (etc) with unit suffix ([pr#30701](#), Prashant D)
- doc: mounting CephFS subdirectory and Persistent Mounts cleanup ([pr#32498](#), Jos Collin)
- doc: Move ceph-deploy docs to doc/install/ceph-deploy ([pr#33953](#), Sebastian Wagner)
- doc: move cephadm files to its own directory ([pr#33551](#), Alexandra Settle, Sebastian Wagner)
- doc: move Developer Guide to its own subdirectory ([pr#27159](#), Nathan Cutler)
- doc: nautilus 14.2.2 release notes, take three ([pr#29171](#), Nathan Cutler)
- doc: Nautilus mailmaps ([pr#27092](#), Abhishek Lekshmanan)
- doc: note explicitly that profile rbd allows blacklisting ([pr#28296](#), Matthew Vernon)
- doc: obsolete entries for allow\_standby\_replay ([pr#31897](#), Rodrigo Severo)
- doc: operations: correct comma-delimited ([pr#29644](#), Anthony DAtri)
- doc: operations: improve reweight-by-utilization ([pr#27657](#), Anthony DAtri)
- doc: PendingReleaseNotes: 14.2.1 note on crush required version ([pr#27649](#), Sage Weil)
- doc: PendingReleaseNotes: fix typo ([pr#31853](#), Sage Weil)
- doc: PendingReleaseNotes: note on python3.6 changes ([issue#39164](#), [pr#27490](#), Kefu Chai)
- doc: pg\_num should always be a power of two ([pr#29364](#), Lars Marowsky-Bree, Kai Wagner)
- doc: QAT Acceleration for Encryption and Compression ([pr#26967](#), Qiaowei Ren)
- doc: quick-rbd.rst de-duplicate ([pr#32965](#), Tim)

- doc: RBD exclusive locks ([pr#31893](#), Florian Haas)
- doc: README.md: remove stale cmake prerequisite ([pr#32751](#), Kefu Chai)
- doc: release note: Add pending release notes for already merged code ([pr#32041](#), David Zafman)
- doc: release notes for 14.2.1 ([pr#27793](#), Abhishek Lekshmanan)
- doc: release notes for Luminous v12.2.13 ([pr#33030](#), Nathan Cutler)
- doc: release notes for nautilus 14.2.2 ([pr#29011](#), Sage Weil, Nathan Cutler)
- doc: release notes for Nautilus 14.2.7 ([pr#33031](#), Nathan Cutler)
- doc: release notes for v14.2.3 nautilus ([pr#29973](#), Abhishek Lekshmanan)
- doc: release notes for v14.2.6 ([pr#32551](#), Abhishek Lekshmanan)
- doc: releases/luminous: release notes for 12.2.12 ([pr#27553](#), Abhishek Lekshmanan)
- doc: releases: 14.2.3 dashboard note ([pr#30145](#), Abhishek Lekshmanan)
- doc: releases: v14.2.8 release notes ([pr#33670](#), Abhishek Lekshmanan)
- doc: relicense LGPL-2.1 code as LGPL-2.1 or LGPL-3.0 ([pr#22446](#), Sage Weil)
- doc: remove prod cluster examples from hardware recs ([pr#32670](#), Zac Dover)
- doc: remove recommendation for kernel.pid\_max ([pr#27965](#), Ben England)
- doc: remove reference to obsolete scrub command ([pr#32508](#), Patrick Donnelly)
- doc: remove the CephFS-Hadoop instructions ([pr#32980](#), Greg Farnum)
- doc: removed OpenStack Kilo references in Keystone docs ([pr#27203](#), James McClune)
- doc: removes kube-helm installation instructions ([pr#32009](#), Zac Dover)
- doc: reorganize CephFS landing page and ToC ([pr#32038](#), Patrick Donnelly)
- doc: Revert doc: do not add suffix for search result links ([pr#33562](#), Jason Dillaman)
- doc: rgw/pubsub: add S3 compliant API to master zone ([pr#28971](#), Yuval Lifshitz)
- doc: rgw/pubsub: clarify pubsub zone configuration ([pr#27493](#), Yuval Lifshitz)
- doc: rgw/pubsub: fix topic arn. tenant support to multisite tests ([pr#27671](#), Yuval Lifshitz)
- doc: rgw: Fixed bug on wrong name for user\_id for OPA ([pr#31972](#), Seena Fallah)
- doc: s/achieve/achieves/ (Fixed a verb disagreement) ([pr#32036](#), Zac Dover)

- doc: script/ceph-backport.sh: add Troubleshooting notes ([pr#29948](#), Nathan Cutler)
- doc: set ceph\_perf\_msgr\_server arguments ([pr#29847](#), Changcheng Liu)
- doc: show how to count jobs before triggering them ([pr#32145](#), Rishabh Dave)
- doc: Show Jenkins commands ([pr#29423](#), Ernesto Puerta)
- doc: Small update of SubmittingPatches-backports ([pr#31163](#), Laura Paduano)
- doc: split up SubmittingPatches.rst ([issue#20953](#), [pr#30705](#), Nathan Cutler)
- doc: Switch spelling of utilization ([pr#32537](#), Bryan Stillwell)
- doc: tools/rados: add -pgid in help ([pr#30383](#), Vikhyat Umrao)
- doc: typo fix in doc/dev/dev\_cluster\_deployment.rst: s/hostanme/hostname/ ([pr#31515](#), Drunkard Zhang)
- doc: update -force flag to be precise ([pr#32343](#), Jos Collin)
- doc: update adding an MDS ([pr#32291](#), Jos Collin)
- doc: update and improve mounting with fuse/kernel docs ([pr#30754](#), Rishabh Dave)
- doc: update bluestore cache settings and clarify data fraction ([issue#39522](#), [pr#27859](#), Jan Fajerski)
- doc: update ceph ansible iscsi info ([pr#28665](#), Mike Christie)
- doc: Update ceph-deploy docs from dumpling to nautilus ([pr#30269](#), Danny Abukalam)
- doc: Update ceph-iscsi min version ([pr#29195](#), Ricardo Marques)
- doc: update CephFS overview in introductory page ([pr#30014](#), Patrick Donnelly)
- doc: update CephFS Quick Start doc ([pr#30406](#), Rishabh Dave)
- doc: Update commands in bootstrap.rst ([pr#31800](#), Zac Dover)
- doc: update default container images ([pr#33974](#), Sage Weil)
- doc: Update documentation for LazyIO methods lazyio\_synchronize() and lazyio\_propagate() ([pr#29711](#), Sidharth Anupkrishnan)
- doc: update documentation for the MANY\_OBJECTS\_PER\_PG warning ([pr#27403](#), Vangelis Tasoulas)
- doc: update documents on using kcephfs ([pr#30626](#), Jeff Layton)
- doc: update erasure-code-profile.rst ([pr#33707](#), Guillaume Abrioux)
- doc: Update link to Red Hat documentation ([pr#27976](#), Yaniv Kaul)

- doc: update list of formats for -format flag for ceph pg dump ([pr#32373](#), Zac Dover)
- doc: Update mailing lists ([pr#31666](#), hrchu)
- doc: update mondb recovery script ([pr#28515](#), Hannes von Haugwitz)
- doc: Update mount CephFS index ([pr#28955](#), Jos Collin)
- doc: Update python-rtsli and tcmu-runner min versions ([pr#28494](#), Ricardo Marques)
- doc: Update requirements for using CephFS ([pr#30251](#), Varsha Rao)
- doc: update with osd addition ([pr#31244](#), Changcheng Liu)
- doc: update with zone bucket and straw2 addition ([pr#31177](#), Changcheng Liu)
- doc: update Zabbix template reference ([pr#33661](#), Mathijs Smit)
- doc: updated ceph monitor config options ([pr#29982](#), James McClune)
- doc: Updated dashboard iSCSI configuration, added labels ([pr#27074](#), Lenz Grimmer)
- doc: updated OpenStack rbd documentation ([pr#28979](#), James McClune)
- doc: updated OS recommendations and distro list ([pr#28643](#), Kai Wagner)
- doc: Updates link to Sepia la ([pr#28780](#), Varsha Rao)
- doc: use subsection for representing components in release notes ([pr#33940](#), Kefu Chai)
- doc: use the console lexer for rendering command line sessions ([pr#32141](#), Kefu Chai)
- do\_cmake.sh: fedora-32 (rawhide) build with python-3.8 ([pr#32474](#), Kaleb S. Keithley)
- errorator: improve general error handlers ([pr#33344](#), Samuel Just)
- github/codeowners: Add orchestrator team ([pr#31441](#), Sebastian Wagner)
- github: Add ceph-volume to list of jenkins commands ([pr#31191](#), Sebastian Wagner)
- include/config-h.in.cmake: remove HAVE\_XIO ([pr#28465](#), Kefu Chai)
- include/utime: do not cast sec to time\_t ([pr#27861](#), Kefu Chai)
- include: buffer\_raw.h: Copyright time fix ([pr#28481](#), Changcheng Liu)
- install-deps.sh: remove failing error catching ([pr#29403](#), Ernesto Puerta)
- Integrate PeeringState into crimson, fix related bugs ([pr#28180](#), Samuel Just)

- krbd: do away with explicit memory management and other cleanups ([pr#31919](#), Ilya Dryomov)
- librados: allow passing flags to operate sync APIs ([pr#33536](#), Yuval Lifshitz)
- librados: fix leak in getxattr and getxattrs ([pr#32183](#), Adam Kupczyk)
- librados: move buffer free functions to inline namespace ([issue#39972](#), [pr#28167](#), Jason Dillaman)
- librados: prefer reinterpret\_cast over c-style cast ([pr#33038](#), Kefu Chai)
- librbd: add reference counting ([pr#30397](#), Mahati Chamarthy, Venky Shankar)
- librbd: add snap\_get\_name and snap\_get\_id method API ([pr#31280](#), Zheng Yin)
- librbd: added missing <string> include to PoolMetadata header ([pr#32614](#), Kaleb S. Keithley)
- librbd: adjust the else-if conditions in validate\_striping() ([pr#30053](#), mxdInspur)
- librbd: always initialize local variables ([pr#31311](#), Kefu Chai)
- librbd: always try to acquire exclusive lock when removing image ([pr#29775](#), Mykola Golub)
- librbd: async open/close should free ImageCtx before issuing callback ([issue#39031](#), [pr#27682](#), Jason Dillaman)
- librbd: avoid dereferencing an empty container during deep-copy ([issue#40368](#), [pr#28559](#), Jason Dillaman)
- librbd: behave more gracefully when data pool removed ([pr#29613](#), Mykola Golub)
- librbd: bump minor version to match octopus ([pr#32402](#), Jason Dillaman)
- librbd: clean up unused variable ([pr#30019](#), mxdInspur)
- librbd: clone copy-on-write operations should preserve sparseness ([pr#27999](#), Mykola Golub)
- librbd: copyup read stats were incorrectly tied to child ([pr#27757](#), Jason Dillaman)
- librbd: defer event socket completion until after callback issued ([pr#33994](#), Jason Dillaman)
- librbd: diff iterate with fast-diff now correctly includes parent ([pr#32403](#), Jason Dillaman)
- librbd: disable zero-copy writes by default ([pr#31794](#), Jason Dillaman)

- librbd: dispatch delayed requests only if read intersects ([pr#27446](#), Mykola Golub)
- librbd: do not allow to deep copy migrating image ([pr#27194](#), Mykola Golub)
- librbd: do not unblock IO prior to growing object map during resize ([issue#39952](#), [pr#28295](#), Jason Dillaman)
- librbd: dont call refresh from mirror::GetInfoRequest state machine ([pr#32734](#), Mykola Golub)
- librbd: dont use complete\_external\_callback if ImageCtx destroyed ([pr#29263](#), Mykola Golub)
- librbd: explicitly specify mode on mirror image enable ([pr#32217](#), Mykola Golub)
- librbd: features converting bitmask and string API ([pr#31188](#), Zheng Yin)
- librbd: finish write request early ([pr#32113](#), Li, Xiaoyan)
- librbd: fix broken group snapshot handling ([pr#33448](#), Jason Dillaman)
- librbd: fix build on freebsd ([pr#32938](#), Mykola Golub)
- librbd: fix issues with object-map/fast-diff feature interlock ([issue#39521](#), [pr#28051](#), Jason Dillaman)
- librbd: fix potential race conditions ([pr#33563](#), Mahati Chamarthy)
- librbd: fix potential snapshot remove failure due to duplicate RPC messages ([pr#32760](#), Mykola Golub)
- librbd: fix rbd\_features\_to\_string output ([pr#31006](#), Zheng Yin)
- librbd: fix rbd\_open\_by\_id, rbd\_open\_by\_id\_read\_only ([pr#32105](#), yangjun)
- librbd: fix some edge cases for snapshot mirror mode promote ([pr#32567](#), Mykola Golub)
- librbd: fix typo in deep\_copy::ObjectCopyRequest::compute\_read\_ops ([pr#27049](#), Mykola Golub)
- librbd: fixed several race conditions related to copyup ([issue#39021](#), [pr#27357](#), Jason Dillaman)
- librbd: force reacquire lock if blacklist is disabled ([pr#30955](#), luo.runbing)
- librbd: implement ordering for overlapping IOs ([pr#28952](#), Mahati Chamarthy)
- librbd: improve journal performance to match expected degradation ([issue#40072](#), [pr#28539](#), Jason Dillaman)

- librbd: improved support for balanced and localized reads ([pr#33493](#), Zheng Yin)
- librbd: initial consolidation of internal locks ([pr#27756](#), Jason Dillaman)
- librbd: introduce new default write-around cache policy ([pr#27229](#), Jason Dillaman)
- librbd: leak on canceling simple io scheduler timer task ([pr#27755](#), Mykola Golub)
- librbd: look for mirror peers in default namespace ([pr#32338](#), Mykola Golub)
- librbd: look for pool metadata in default namespace ([pr#27151](#), Mykola Golub)
- librbd: make flush be queued by QOS throttler ([pr#26931](#), Mykola Golub)
- librbd: mirror image enable/disable should enable/disable journaling ([pr#28553](#), Mykola Golub)
- librbd: optimize image copy state machine to use fast-diff ([pr#33867](#), Jason Dillaman)
- librbd: optionally move parent image to trash on remove ([pr#27521](#), Mykola Golub)
- librbd: prevent concurrent AIO callbacks to external clients ([issue#40417](#), [pr#28743](#), Jason Dillaman)
- librbd: Remove duplicated AsyncOpTracker in librbd/Utils.h ([pr#29653](#), Xiaoyan Li)
- librbd: remove pool objects when removing a namespace ([pr#32401](#), Jason Dillaman)
- librbd: shared read-only cache hook ([pr#27285](#), Dehao Shang, Yuan Zhou)
- librbd: silence -Wunused-variable warnings ([pr#27513](#), David Disseldorp)
- librbd: simple scheduler plugin for object dispatcher layer ([pr#26675](#), Mykola Golub)
- librbd: snapshot object maps can go inconsistent during copyup ([issue#39435](#), [pr#27724](#), Ilya Dryomov)
- librbd: support compression allocation hints to the OSD ([pr#32687](#), Jason Dillaman)

- librbd: support EC data pool images sparsify ([pr#27268](#), Mykola Golub)
- librbd: support zero-copy writes via the C API ([pr#27895](#), Jason Dillaman)
- librbd: trash move return EBUSY instead of EINVAL for migrating image ([pr#27136](#), Mykola Golub)
- librbd: tweak deep-copy to avoid creating last snapshot until sync is complete ([pr#33097](#), Jason Dillaman)
- librbd: tweaks to increase IOPS and reduce CPU usage ([pr#28044](#), Jason Dillaman)
- librbd: use custom allocator for aligned boost::lockfree::queue ([issue#39703](#), [pr#28093](#), Jason Dillaman)
- librbd: v1 clones are restricted to the same namespace ([pr#30711](#), Jason Dillaman)
- librbd: when unlinking peer from mirror snaps do it in all namespaces ([pr#32463](#), Mykola Golub)
- librbd:move all snapshot API functions in internal.cc over to api/Snapshot.cc ([pr#31589](#), Zheng Yin)
- log: avoid logging anything when log\_to\_file=false ([pr#27133](#), Sage Weil)
- log: fix store\_statfs log line ([pr#28564](#), Mohamad Gebai)
- log: just return if t is empty ([pr#31243](#), Xiubo Li)
- log: print pthread ID / name mapping in recent events dump ([pr#32354](#), Radoslaw Zarzynski)
- lvm deactivate command ([pr#32179](#), Jan Fajerski)
- mds: add command that config individual client session ([issue#40811](#), [pr#29104](#), Yan, Zheng)
- mds: add config to require forward to auth MDS ([pr#29995](#), simon gao)
- mds: add configurable snapshot limit ([pr#30710](#), Milind Changire)
- mds: add perf counter for finisher of MDSRank ([pr#29377](#), simon gao)
- mds: add perf counters for openfiletable ([pr#33363](#), Milind Changire)
- mds: add scrub\_info\_t into mempool ([pr#33180](#), Jun Su)
- mds: answering all pending getattr/lookups targeting the same inode in one go ([issue#36608](#), [pr#24794](#), Patrick Donnelly, Xuehan Xu)
- mds: apply configuration changes through MDSRank ([pr#28951](#), Patrick Donnelly)

- mds: async dir operation support ([pr#27866](#), Yan, Zheng)
- mds: async dirop support ([pr#32816](#), Yan, Zheng)
- mds: avoid check session connections features when issuing caps ([pr#26881](#), Yan, Zheng)
- mds: avoid revoking Fsx from loner during directory fragmentation ([pr#26817](#), Yan, Zheng)
- mds: avoid sending too many osd requests at once after mds restarts ([issue#40028](#), [pr#27436](#), simon gao)
- mds: better output of ceph health detail when some client is failing to advance oldest client/flush tid ([issue#39266](#), [pr#27537](#), Shen Hang)
- mds: check dir fragment to split dir if mkdir makes it oversized ([pr#27480](#), Erqi Chen)
- mds: check directory split after rename ([issue#38994](#), [pr#27214](#), Shen Hang)
- mds: clarify comment ([pr#31401](#), Patrick Donnelly)
- mds: cleanup truncating inodes when standby replay mds trim log segments ([pr#28686](#), Yan, Zheng)
- mds: cleanup unneeded client\_snap\_caps when splitting snap inode ([issue#39987](#), [pr#28190](#), Yan, Zheng)
- mds: complete all the replay op when mds is restarted ([issue#40784](#), [pr#29059](#), Shen Hang)
- mds: convert unnecessary usage of std::list to std::vector ([pr#26895](#), Patrick Donnelly)
- mds: count purge queue items left in journal ([issue#40121](#), [pr#28376](#), Zhi Zhang)
- mds: delay exporting directory whose pin value exceeds max rank id ([issue#40603](#), [pr#28804](#), Zhi Zhang)
- mds: display scrub status in ceph status ([pr#28855](#), Venky Shankar)
- mds: do not include metric\_spec in MClientSession from MDS ([pr#32659](#), Patrick Donnelly)
- mds: dont add metadata to session close message ([pr#32318](#), Yan, Zheng)
- mds: dont mark cap NEEDSNAPFLUSH if client has no pending capsnap ([pr#28551](#), Yan, Zheng)
- mds: dont print subtrees if they are too big or too many ([pr#26056](#), Rishabh Dave)

- mds: dont respond getattr with -EROFS when mds is readonly ([pr#32676](#), Yan, Zheng)
- mds: drive cap recall while dropping cache ([pr#30389](#), Patrick Donnelly)
- mds: evict an unresponsive client only when another client wants its caps ([issue#17854](#), [pr#22645](#), Rishabh Dave)
- mds: execute PurgeQueue on\_error handler in finisher ([pr#29064](#), Yan, Zheng)
- mds: fix assert(omap\_num\_objs <= MAX\_OBJECTS) of OpenFileTable ([pr#32020](#), Yan, Zheng)
- mds: fix bug of batch getattr/lookup ([pr#32268](#), Yan, Zheng)
- mds: fix can wrlock check in Locker::acquire\_locks() ([pr#33005](#), Yan, Zheng)
- mds: fix infinite loop in Locker::file\_update\_finish ([pr#29902](#), Yan, Zheng)
- mds: fix InoTable::force\_consume\_to() ([pr#29411](#), Yan, Zheng)
- mds: fix invalid access of mdr->dn[0].back() ([pr#31534](#), Yan, Zheng)
- mds: fix is session in blacklist check in Server::apply\_blacklist() ([issue#40061](#), [pr#28293](#), Yan, Zheng)
- mds: Fix MDCache.h reorder compiler warnings ([pr#31409](#), Varsha Rao)
- mds: fix null pointer dereference in Server::handle\_client\_link() ([pr#32722](#), Yan, Zheng)
- mds: fix revoking caps after after stale->resume circle ([pr#31662](#), Yan, Zheng)
- mds: fix SnapRealm::resolve\_snapname for long name ([pr#27511](#), Yan, Zheng)
- mds: fix use-after-free in Migrater ([pr#33291](#), Yan, Zheng)
- mds: handle bad purge queue item encoding ([pr#33449](#), Yan, Zheng)
- mds: handle ceph\_assert on blacklisting ([pr#33662](#), Milind Changire)
- mds: increase default cache memory limit to 4G ([pr#32042](#), Patrick Donnelly)
- mds: initialize cap\_revoke\_eviction\_timeout with conf ([issue#38844](#), [pr#26970](#), simon gao)
- mds: initialize the monc later in init() ([pr#31715](#), Xiubo Li)
- mds: just delete MDSIOContextBase during shutdown ([pr#33538](#), Patrick Donnelly)
- mds: maintain client provided metric flags in client metadata ([pr#32201](#), Venky Shankar)
- mds: make mds-mds per-message versioned ([issue#12107](#), [pr#20160](#), dongdong tao)

- mds: make MDSIOContextBase delete itself when shutting down ([pr#29752](#), Xuehan Xu)
- mds: mds returns -5(EIO) error when the deleted file does not exist ([pr#30403](#), huanwen ren)
- mds: move some MDCache member init to header ([pr#29543](#), Patrick Donnelly)
- mds: no assert on frozen dir when scrub path ([pr#30835](#), Zhi Zhang)
- mds: note client features when rejecting client ([pr#32505](#), Patrick Donnelly)
- mds: obsoleting mds\_cache\_size ([pr#31729](#), Patrick Donnelly, Ramana Raja)
- mds: optimize function, fragset\_t::simplify, to improve the efficiency of merging fragment ([pr#31595](#), simon gao)
- mds: output lock state in format dump ([issue#39645](#), [pr#27717](#), Zhi Zhang)
- mds: pass proper MutationImpl::LockOp to Locker::wrlock\_start() ([pr#33719](#), Yan, Zheng)
- mds: preparation for async dir operation support ([pr#30972](#), Yan, Zheng)
- mds: properly evaluate unstable locks when evicting client ([pr#31548](#), Yan, Zheng)
- mds: recall caps from quiescent sessions ([pr#28702](#), Patrick Donnelly)
- mds: register with mgr only after added to FSMMap ([pr#31400](#), Patrick Donnelly)
- mds: reject sessionless messages ([pr#29594](#), Xiao Guodong)
- mds: release free heap pages after trim ([pr#31793](#), Patrick Donnelly)
- mds: relevel debug message levels for balancer/migrator ([pr#33471](#), Patrick Donnelly)
- mds: remove dead get\_commands code ([pr#33390](#), Patrick Donnelly)
- mds: remove duplicated check on balance amount ([pr#27087](#), Zhi Zhang)
- mds: remove superfluous error in StrayManager::advance\_delayed() ([issue#38679](#), [pr#27051](#), Yan, Zheng)
- mds: remove the code that skip evicting the only client ([pr#28642](#), Yan, Zheng)
- mds: remove the incorrect comments ([pr#31775](#), Xiubo Li)
- mds: remove unnecessary debug warning ([pr#31898](#), Patrick Donnelly)
- mds: remove unused CDir members ([pr#33227](#), Jun Su)
- mds: Reorganize class members in Anchor header ([pr#30090](#), Varsha Rao)

- mds: Reorganize class members in Capability header ([pr#29166](#), Varsha Rao)
- mds: Reorganize class members in CDir header ([pr#28860](#), Varsha Rao)
- mds: Reorganize class members in CIinode header ([pr#29066](#), Varsha Rao)
- mds: Reorganize class members in DamageTable header ([pr#29569](#), Varsha Rao)
- mds: Reorganize class members in FSMap header ([pr#29572](#), Varsha Rao)
- mds: Reorganize class members in FSMapUser header ([pr#29574](#), Varsha Rao)
- mds: Reorganize class members in InoTable header ([pr#29883](#), Varsha Rao)
- mds: Reorganize class members in JournalPointer header ([pr#29888](#), Varsha Rao)
- mds: Reorganize class members in LocalLock header ([pr#30143](#), Varsha Rao)
- mds: Reorganize class members in Locker header ([pr#30164](#), Varsha Rao)
- mds: Reorganize class members in LogEvent header ([pr#30205](#), Varsha Rao)
- mds: Reorganize class members in LogSegment header ([pr#30202](#), Varsha Rao)
- mds: Reorganize class members in MDBalancer header ([pr#30559](#), Varsha Rao)
- mds: Reorganize class members in MDCache header ([pr#30745](#), Varsha Rao)
- mds: Reorganize class members in MDLog header ([pr#30744](#), Varsha Rao)
- mds: Reorganize class members in MDSAuthCaps header ([pr#30915](#), Varsha Rao)
- mds: Reorganize class members in MDSCacheObject header ([pr#30938](#), Varsha Rao)
- mds: Reorganize class members in MDSDaemon header ([pr#30990](#), Varsha Rao)
- mds: Reorganize class members in MDSMap header ([pr#31118](#), Varsha Rao)
- mds: Reorganize class members in MDSRank header ([pr#31120](#), Varsha Rao)
- mds: Reorganize class members in MDSTable header ([pr#31122](#), Varsha Rao)
- mds: Reorganize class members in MDSTableClient header ([pr#31115](#), Varsha Rao)
- mds: Reorganize class members in MDSTableServer header ([pr#31250](#), Varsha Rao)
- mds: Reorganize class members in Migrator header ([pr#31253](#), Varsha Rao)
- mds: Reorganize class members in OpenFileTable header ([pr#31597](#), Varsha Rao)
- mds: Reorganize class members in PurgeQueue header ([pr#31596](#), Varsha Rao)
- mds: Reorganize class members in RecoveryQueue header ([pr#31635](#), Varsha Rao)

- mds: Reorganize class members in ScatterLock header ([pr#31716](#), Varsha Rao)
- mds: Reorganize class members in ScrubHeader header ([pr#31717](#), Varsha Rao)
- mds: Reorganize class members in ScrubStack header ([pr#31718](#), Varsha Rao)
- mds: Reorganize class members in Server header ([pr#31719](#), Varsha Rao)
- mds: Reorganize class members in SessionMap header ([pr#32320](#), Varsha Rao)
- mds: Reorganize class members in SimpleLock header ([pr#32322](#), Varsha Rao)
- mds: Reorganize class members in SnapClient header ([pr#32326](#), Varsha Rao)
- mds: Reorganize class members in SnapServer header ([pr#32350](#), Varsha Rao)
- mds: Reorganize struct members in Mutation header ([pr#31481](#), Varsha Rao)
- mds: Reorganize structure and class members in mdstypes header ([pr#32435](#), Varsha Rao)
- mds: Reorganize structure members in flock header ([pr#32416](#), Varsha Rao)
- mds: Reorganize structure members in inode\_backtrace header ([pr#32431](#), Varsha Rao)
- mds: Reorganize structure members in snap header ([pr#32432](#), Varsha Rao)
- mds: Reorganize structure members in SnapRealm header ([pr#32348](#), Varsha Rao)
- mds: Reorganize structure members in StrayManager header ([pr#32397](#), Varsha Rao)
- mds: reset heartbeat inside big loop ([pr#28406](#), Yan, Zheng)
- mds: split the dir if the op makes it oversized, because some ops maybe in flight ([pr#29921](#), simon gao)
- mds: there is an assertion when calling Beacon::shutdown() ([issue#38822](#), [pr#27063](#), huanwen ren)
- mds: throttle scrub start for multiple active MDS ([pr#32521](#), Patrick Donnelly, Milind Changire)
- mds: tolerate no snaprealm encoded in on-disk root inode ([pr#31455](#), Yan, Zheng)
- mds: track high water mark for purges ([pr#32667](#), Patrick Donnelly)
- mds: trim cache during standby-replay ([issue#40213](#), [pr#28212](#), simon gao)
- mds: trim cache on regular schedule ([pr#29542](#), Patrick Donnelly)
- mds: unify daemon and tell commands ([pr#31255](#), Sage Weil)

- mds: update projected\_version when upgrading snaptable ([issue#38835](#), [pr#27238](#), Yan, Zheng)
- mds: use set to store to evict client ([pr#30029](#), Erqi Chen)
- mds: use vector::empty in feature\_bitset\_t ([pr#32541](#), Jos Collin)
- mds: wake up lock waiters after forcibly changing lock state ([issue#39987](#), [pr#28459](#), Yan, Zheng)
- mgr,mon,rbd: mon/mgr: add rbd\_support to list of always-on mgr modules ([issue#40790](#), [pr#29073](#), Jason Dillaman)
- mgr,mon: mon,mgr: pass MessageRef to monc.send\_mon\_message() xe2x80xa6 ([pr#30449](#), Kefu Chai)
- mgr,mon: mon/MgrMonitor.cc: add always\_on\_modules to the output of ceph mgr module ls ([pr#32939](#), Neha Ojha)
- mgr,mon: mon/MgrMonitor.cc: warn about missing mgr in a cluster with osds ([pr#33025](#), Neha Ojha)
- mgr,pybind: pybind/mgr/prometheus: remove scrape\_duration metric ([pr#27034](#), Jan Fajerski)
- mgr,rbd: mgr/dashboard: block mirroring page results in internal server error ([pr#31907](#), Jason Dillaman)
- mgr,rbd: mgr/rbd\_support: dont scan pools that dont have schedules ([pr#33840](#), Mykola Golub)
- mgr,rbd: mgr/rbd\_support: implement mirror snapshot scheduler ([pr#32434](#), Mykola Golub)
- mgr,rbd: mgr/rbd\_support: support scheduling long-running background operations ([issue#40621](#), [pr#29054](#), Jason Dillaman)
- mgr,rbd: pybind/mgr: fix format for rbd-mirror prometheus metrics ([pr#28200](#), Mykola Golub)
- mgr,rgw: mgr/ansible: RGW service ([pr#28468](#), Juan Miguel Olmo Martxc3xadnez)
- mgr,tests: install-deps.sh: preload wheel for all mgr requirements.txt files ([pr#32151](#), Sage Weil)
- mgr,tests: mgr/orchestrator\_cli: remove tox and move test to parent dir ([pr#31561](#), Sebastian Wagner)
- mgr,tests: mgr/progress: Created first unit test for progress module ([pr#28758](#), Kamoltat (Junior) Sirivadhna)
- mgr,tests: pybind/mgr: Add ceph\_module.pyi to improve type checking ([pr#32502](#),

Sebastian Wagner)

- mgr, tests: pybind/mgr: install setuptools >= 12 ([pr#29414](#), Kefu Chai)
- mgr, tests: pybind/tox: handle possible WITH\_PYTHON3 values other than 3 ([pr#28002](#), Nathan Cutler)
- mgr, tests: qa/mgr/balancer: Add cram based test for altering target\_max\_misplaced\_ratio setting ([pr#30646](#), Shyukri Shyukriev)
- mgr, tests: qa/mgr/progress: update the test suite for progress module ([issue#40618](#), [pr#29111](#), Kamoltat (Junior) Sirivadhna)
- mgr/tools: Remove use of rules batching for upmap balancer and default for upmap\_max\_deviation to 5 ([pr#32247](#), David Zafman)
- mgr/ansible: Host ls implementation ([pr#26185](#), Juan Miguel Olmo Martxc3xadnez)
- mgr/ansible: Integrate mgr/ansible/tox into mgr/tox ([pr#32149](#), Sebastian Wagner)
- mgr/ansible: TLS Mutual Authentication ([pr#27512](#), Juan Miguel Olmo Martxc3xadnez)
- mgr/cephadm: a few fixes around daemon and device caches ([pr#33495](#), Sage Weil)
- mgr/cephadm: adapt osd deployment to service\_apply ([pr#33922](#), Sage Weil, Joshua Schmid)
- mgr/cephadm: add drivegroup support; workaround c-v batch shortcoming ([pr#32972](#), Sage Weil, Joshua Schmid)
- mgr/cephadm: add HostAssignment.validate() ([pr#34005](#), Sebastian Wagner)
- mgr/cephadm: Add progress to update\_mgr() ([pr#32372](#), Sebastian Wagner)
- mgr/cephadm: Add unittest for osd removal ([pr#33602](#), Sage Weil, Sebastian Wagner)
- mgr/cephadm: Add unittest for service\_action ([pr#32209](#), Sebastian Wagner)
- mgr/cephadm: allow osd replacement/removal in the background ([pr#32983](#), Joshua Schmid)
- mgr/cephadm: auto-select python version to use remotely ([pr#32327](#), Sage Weil)
- mgr/cephadm: cache device inventory; zap ([pr#33394](#), Sage Weil)
- mgr/cephadm: catch exceptions when scraping ceph-volume inventory ([pr#33484](#), Sage Weil)
- mgr/cephadm: catch exceptions in serve() thread ([pr#33139](#), Sage Weil)
- mgr/cephadm: check-host on host add ([pr#32385](#), Sage Weil)
- mgr/cephadm: clean up client.crash.\\* container\_image settings after upgrade

([pr#34068](#), Sage Weil)

- mgr/cephadm: consolidate/refactor all add\\_ and apply\\_ methods ([pr#33496](#), Sage Weil)
- mgr/cephadm: Convert HostNotFound to OrchestratorError ([pr#33310](#), Sebastian Wagner)
- mgr/cephadm: deploy Grafana ([pr#33515](#), Patrick Seidensal)
- mgr/cephadm: do not include osd service in orch ls output ([pr#33968](#), Sage Weil)
- mgr/cephadm: do not reconfig orphan daemons; fix test to not remote orphans ([pr#34027](#), Sage Weil)
- mgr/cephadm: do not refresh daemon and device inventory as often ([pr#33734](#), Sage Weil)
- mgr/cephadm: drop mixin parent ([pr#33514](#), Sage Weil)
- mgr/cephadm: Enable provisioning alertmanager via orchestrator ([pr#33554](#), Kristoffer Grxc3xb6nlund)
- mgr/cephadm: fix dump output by formatting to yaml first ([pr#33891](#), Joshua Schmid)
- mgr/cephadm: fix listing services by host ([pr#32314](#), Kiefer Chang)
- mgr/cephadm: fix orch rm and upgrade ([pr#33772](#), Sage Weil)
- mgr/cephadm: fix osd reconfig/redeploy ([pr#32812](#), Sage Weil)
- mgr/cephadm: Fix placement for new services ([pr#33205](#), Sebastian Wagner)
- mgr/cephadm: fix placement when existing + specified dont overlap ([pr#33766](#), Sage Weil)
- mgr/cephadm: fix prom config generation when hosts have no labels or addrs ([pr#33800](#), Sage Weil)
- mgr/cephadm: Fix remove\_osds() ([pr#32146](#), Sebastian Wagner)
- mgr/cephadm: fix section name for mon options in ceph.conf ([pr#32681](#), Sage Weil)
- mgr/cephadm: fix service list filtering ([pr#33838](#), Kiefer Chang)
- mgr/cephadm: fix type of timeout options ([pr#32316](#), Kiefer Chang)
- mgr/cephadm: fix upgrade ok-to-stop condition check ([pr#33469](#), Sage Weil)
- mgr/cephadm: fix upgrade order ([pr#33811](#), Sage Weil)
- mgr/cephadm: fix upgrade wait loop ([pr#33447](#), Sage Weil)

- mgr/cephadm: fix upgrade when daemon is stopped ([pr#33678](#), Sage Weil)
- mgr/cephadm: if we had no record of deps, and deps are [], do not reconfig ([pr#33733](#), Sage Weil)
- mgr/cephadm: implement apply mon, mon removal checks ([pr#33792](#), Sage Weil)
- mgr/cephadm: implement pause/resume to suspect non-monitoring background work ([pr#33930](#), Sage Weil)
- mgr/cephadm: improve pull behavior for upgrade ([pr#32878](#), Sage Weil)
- mgr/cephadm: init attrs created by setattr() ([pr#32957](#), Kefu Chai)
- mgr/cephadm: leverage service specs ([pr#33553](#), Sage Weil, Joshua Schmid)
- mgr/cephadm: limit number of times check host is performed in the serve loop ([pr#33866](#), Daniel-Pivonka)
- mgr/cephadm: log information to cluster log ([pr#33488](#), Sage Weil)
- mgr/cephadm: make apply move daemons, do its work synchronously ([pr#33704](#), Sage Weil)
- mgr/cephadm: make NodeAssignment return a simple host list ([pr#33669](#), Sage Weil)
- mgr/cephadm: make osd create on an existing LV idempotent ([pr#33755](#), Sage Weil)
- mgr/cephadm: make prometheus scrape all mgrs, node-exporters ([pr#33444](#), Sage Weil)
- mgr/cephadm: Make sure we dont co-locate the same daemon ([pr#33853](#), Sebastian Wagner)
- mgr/cephadm: misc fixes ([pr#33119](#), Sage Weil)
- mgr/cephadm: misc fixes + smoke test ([pr#33730](#), Sage Weil)
- mgr/cephadm: mon: Dont show traceback for user errors ([pr#33333](#), Sebastian Wagner)
- mgr/cephadm: nicer error from cephadm check-host ([pr#33935](#), Sage Weil)
- mgr/cephadm: point dashboard at cephadms grafana automatically ([pr#33700](#), Sage Weil)
- mgr/cephadm: prefix daemon ids with hostname ([pr#33012](#), Sage Weil)
- mgr/cephadm: progress for upgrade ([pr#33415](#), Sage Weil)
- mgr/cephadm: provision node-exporter ([pr#33123](#), Sage Weil, Patrick Seidensal)
- mgr/cephadm: provision prometheus ([pr#33073](#), Sage Weil)

- mgr/cephadm: reduce boilerplate for unittests ([pr#33663](#), Joshua Schmid)
- mgr/cephadm: refresh ceph.conf when mons change ([pr#33855](#), Sage Weil)
- mgr/cephadm: refresh configs when dependencies change ([pr#33671](#), Sage Weil)
- mgr/cephadm: refresh service state in the background ([pr#32859](#), Sebastian Wagner, Sage Weil)
- mgr/cephadm: remove item from cache when removing ([pr#33071](#), Sage Weil)
- mgr/cephadm: remove redundant /dev when blinking device light ([pr#32246](#), Sage Weil)
- mgr/cephadm: revamp scheduling ([pr#33523](#), Sage Weil)
- mgr/cephadm: set thread pool size to 10 ([pr#33463](#), Sebastian Wagner)
- mgr/cephadm: show age of service ls ([pr#32686](#), Sage Weil)
- mgr/cephadm: simplify and improve placement ([pr#33808](#), Sage Weil)
- mgr/cephadm: simplify tracking of daemon inventory ([pr#33249](#), Sage Weil)
- mgr/cephadm: two minor fixes ([pr#33736](#), Sage Weil)
- mgr/cephadm: update osd removal report immediately ([pr#33713](#), Kiefer Chang)
- mgr/cephadm: update type annotation ([pr#33784](#), Kefu Chai)
- mgr/cephadm: upgrade requires root mode for now ([pr#33802](#), Sage Weil)
- mgr/cephadm: upgrade: fix daemons missing image\_id ([pr#33745](#), Sage Weil)
- mgr/cephadm: upgrade: handle stopped daemons ([pr#33487](#), Sage Weil)
- mgr/cephadm: verify hosts hostname matches cephadm host ([pr#33058](#), Sage Weil)
- mgr/dashbaord: Fix E2E pools page failure ([pr#32635](#), Stephan Mxc3xbcller)
- mgr/dashboard: Improve iSCSI overview page ([pr#27254](#), Ricardo Marques)
- mgr/dashboard Displays progress bar in notification tray for background tasks ([pr#27420](#), Pooja)
- mgr/dashboard/qa: Improve tasks.mgr.test\_dashboard.TestDashboard.test\_standby ([pr#26925](#), Volker Theile)
- mgr/dashboard/qa: Increase timeout for test\_disable  
(tasks.mgr.dashboard.test\_mgr\_module.MgrModuleTelemetryTest) ([pr#27187](#), Volker Theile)
- mgr/dashboard: 1 osds exist in the crush map but not in the osdmap breaks OSD

page ([issue#36086](#), [pr#26836](#), Patrick Nawracay)

- mgr/dashboard: A block-manager can not access the pool page ([pr#30001](#), Volker Theile)
- mgr/dashboard: accept expected exception when SSL handshaking ([pr#31014](#), Kefu Chai)
- mgr/dashboard: Access control database does not restore disabled users correctly ([pr#29614](#), Volker Theile)
- mgr/dashboard: adapt bucket tenant API tests to new behaviour ([pr#29570](#), alfonsomthd)
- mgr/dashboard: adapt create\_osds interface change ([pr#34000](#), Kiefer Chang)
- mgr/dashboard: Add Always-on column to mgr module list ([pr#33429](#), Volker Theile)
- mgr/dashboard: Add date range and log search functionality ([issue#37387](#), [pr#26562](#), guodan1)
- mgr/dashboard: add debug mode ([pr#30522](#), Ernesto Puerta)
- mgr/dashboard: add feature toggle for NFS and fix feature toggles regression ([pr#32419](#), Ernesto Puerta)
- mgr/dashboard: Add invalid pattern message for Pool name ([pr#31607](#), Tiago Melo)
- mgr/dashboard: Add missing text translation ([pr#29934](#), Volker Theile)
- mgr/dashboard: Add polish translation ([pr#27247](#), Sebastian Krah)
- mgr/dashboard: Add protractor-screenshoter-plugin ([pr#27166](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Add refresh interval to the dashboard landing page ([issue#26872](#), [pr#26396](#), guodan1)
- mgr/dashboard: Add separate option to config SSL port ([pr#26914](#), Volker Theile)
- mgr/dashboard: Add support for blinking enclosure LEDs ([pr#31851](#), Volker Theile)
- mgr/dashboard: Add time-diff unittest and docs ([pr#31357](#), Volker Theile)
- mgr/dashboard: Add vertical menu ([pr#31923](#), Tiago Melo)
- mgr/dashboard: Add whitelist to guard ([pr#27406](#), Ernesto Puerta)
- mgr/dashboard: Allow deletion of RBD with snapshots ([pr#33067](#), Tiago Melo)
- mgr/dashboard: Allow disabling redirection on standby Dashboards ([pr#29088](#), Volker Theile)

- mgr/dashboard: allow refreshing inventory page ([pr#32423](#), Kiefer Chang)
- mgr/dashboard: Allow users to change their password on the UI ([pr#28935](#), Volker Theile)
- mgr/dashboard: auth ttl expired error ([pr#27098](#), ming416)
- mgr/dashboard: Back button component ([pr#27164](#), Stephan Mxc3xbcller)
- mgr/dashboard: behave when pwdUpdateRequired key is missing ([pr#33513](#), Sage Weil)
- mgr/dashboard: Bucket names cannot be formatted as IP address ([pr#30620](#), Volker Theile)
- mgr/dashboard: ceph dashboard i18ntool ([pr#26953](#), Sebastian Krah)
- mgr/dashboard: CephFS client tab switch ([pr#29556](#), Stephan Mxc3xbcller)
- mgr/dashboard: CephFS tab component ([pr#29800](#), Stephan Mxc3xbcller)
- mgr/dashboard: Change the provider of services to root ([issue#39996](#), [pr#28211](#), Tiago Melo)
- mgr/dashboard: change warn\_explicit to warn ([pr#30075](#), Ernesto Puerta)
- mgr/dashboard: Check if gateway is in use before deletion ([pr#27262](#), Ricardo Marques)
- mgr/dashboard: Check if num\_sessions is available ([pr#30270](#), Ricardo Marques)
- mgr/dashboard: cheroot moved into a separate project ([pr#31431](#), Joshua Schmid)
- mgr/dashboard: Cleanup code ([pr#33107](#), Volker Theile)
- mgr/dashboard: Cleanup feature toggle status output ([pr#32569](#), Volker Theile)
- mgr/dashboard: Cleanup Python code ([pr#29604](#), Volker Theile)
- mgr/dashboard: Clone an existing user role ([pr#32653](#), Volker Theile)
- mgr/dashboard: commands to set SSL certificate and key ([pr#27463](#), Ricardo Dias)
- mgr/dashboard: Configuring an URL prefix does not work as expected ([pr#30599](#), Volker Theile)
- mgr/dashboard: consider mon\_allow\_pool\_delete flag ([pr#28260](#), Tatjana Dehler)
- mgr/dashboard: Controls UI inputs based on type ([pr#30208](#), Ricardo Marques)
- mgr/dashboard: coverage venv python version same as mgr ([pr#33407](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Create bucket with x-amz-bucket-object-lock-enabled ([pr#33821](#),

Volker Theile)

- mgr/dashboard: Crush rule modal ([pr#33620](#), Stephan Mxc3xbcller)
- mgr/dashboard: decouple backend unit tests from build ([pr#32565](#), Alfonso Martxc3xadnez)
- mgr/dashboard: destroyed view in CRUSH map viewer ([pr#33405](#), Avan Thakkar)
- mgr/dashboard: Disable event propagation in the helper icon ([issue#40715](#), [pr#29105](#), Tiago Melo)
- mgr/dashboard: Display correct dialog title ([pr#28168](#), Volker Theile)
- mgr/dashboard: Display iSCSI logged in info ([pr#28265](#), Ricardo Marques)
- mgr/dashboard: Display legend for CephFS standbys ([pr#29927](#), Volker Theile)
- mgr/dashboard: display OSD IDs on inventory page ([pr#31189](#), Kiefer Chang)
- mgr/dashboard: Display the number of iSCSI active sessions ([pr#27248](#), Ricardo Marques)
- mgr/dashboard: Display WWN and LUN number in iSCSI target details ([pr#30288](#), Ricardo Marques)
- mgr/dashboard: do not log tokens ([pr#30445](#), Kefu Chai)
- mgr/dashboard: do not show RGW API keys if only read-only privileges ([pr#33178](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Editing RGW bucket fails because of name is already in use ([pr#29767](#), Volker Theile)
- mgr/dashboard: Enable compiler options used by Angular -strict flag ([pr#32553](#), Tiago Melo)
- mgr/dashboard: Enable read only users to read again ([pr#27348](#), Stephan Mxc3xbcller)
- mgr/dashboard: enable/disable versioning on RGW bucket ([pr#29460](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Enforce password change upon first login ([pr#32680](#), Volker Theile, Tatjana Dehler)
- mgr/dashboard: Enhance user create CLI command to force password change ([pr#33552](#), Volker Theile)
- mgr/dashboard: Evict a CephFS client ([pr#28898](#), Ricardo Marques)
- mgr/dashboard: Explicitly set/change the device class of an OSD ([pr#32150](#),

Ricardo Marques)

- mgr/dashboard: Extend E2E test section ([pr#28858](#), Laura Paduano)
- mgr/dashboard: extend types of smart response ([pr#30595](#), Patrick Seidensal)
- mgr/dashboard: fix adding/removing host errors ([pr#34023](#), Kiefer Chang)
- mgr/dashboard: fix backend error when updating RBD interlocked features ([issue#39933](#), [pr#28147](#), Kiefer Chang)
- mgr/dashboard: fix cdEncode decorator is not working on class ([pr#30064](#), Kiefer Chang)
- mgr/dashboard: Fix CephFS chart ([pr#29557](#), Stephan Mxc3xbcller)
- mgr/dashboard: Fix dashboard health test failure ([pr#29172](#), Ricardo Marques)
- mgr/dashboard: Fix deletion of NFS protocol properties ([issue#38997](#), [pr#27244](#), Tiago Melo)
- mgr/dashboard: Fix deletion of NFS transports properties ([issue#39090](#), [pr#27350](#), Tiago Melo)
- mgr/dashboard: Fix e2e chromedriver problem ([pr#32224](#), Tiago Melo)
- mgr/dashboard: Fix env vars of run-tox.sh ([issue#38798](#), [pr#26977](#), Patrick Nawracay)
- mgr/dashboard: Fix error in unit test caused by timezone ([pr#31632](#), Tiago Melo)
- mgr/dashboard: fix failing user test ([pr#32461](#), Tatjana Dehler)
- mgr/dashboard: fix improper URL checking ([pr#32652](#), Ernesto Puerta)
- mgr/dashboard: Fix iSCSI + Rook issues ([issue#39586](#), [pr#26341](#), Sebastian Wagner)
- mgr/dashboard: Fix iSCSI Discovery user permissions ([issue#39328](#), [pr#27678](#), Tiago Melo)
- mgr/dashboard: Fix iSCSI disk diff calculation ([pr#27378](#), Ricardo Marques)
- mgr/dashboard: Fix iSCSI form when using IPv6 ([pr#27946](#), Ricardo Marques)
- mgr/dashboard: Fix iSCSI target form warning ([issue#39324](#), [pr#27609](#), Tiago Melo)
- mgr/dashboard: Fix iSCSI target submission ([pr#27380](#), Ricardo Marques)
- mgr/dashboard: Fix issues in user form ([pr#28863](#), Volker Theile)
- mgr/dashboard: fix LazyUUID4 not serializable ([pr#31266](#), Ernesto Puerta)
- mgr/dashboard: fix MDS counter chart is not displayed ([pr#29371](#), Kiefer Chang)

- mgr/dashboard: fix mgr module API tests ([pr#29634](#), alfonsomthd, Kefu Chai)
- mgr/dashboard: fix missing constraints file in backend API tests ([pr#30720](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Fix missing i18n ([pr#32650](#), Volker Theile)
- mgr/dashboard: Fix mypy issues and enable it by default ([pr#33454](#), Volker Theile)
- mgr/dashboard: Fix NFS pseudo validation ([issue#39063](#), [pr#27293](#), Tiago Melo)
- mgr/dashboard: Fix NFS squash default value ([issue#39064](#), [pr#27294](#), Tiago Melo)
- mgr/dashboard: Fix npm vulnerabilities ([pr#32699](#), Tiago Melo)
- mgr/dashboard: Fix OSD IDs are not displayed when using cephadm backend ([pr#32207](#), Kiefer Chang)
- mgr/dashboard: Fix pool deletion e2e ([pr#29993](#), Volker Theile)
- mgr/dashboard: Fix pool renaming functionality ([pr#31617](#), Stephan Mxc3xbcller)
- mgr/dashboard: fix python2 failure in home controller ([pr#30937](#), Ricardo Dias)
- mgr/dashboard: fix RGW subuser auto-generate key ([pr#32186](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Fix RGW user/bucket quota issues ([pr#28174](#), Volker Theile)
- mgr/dashboard: fix SAML input argument handling ([pr#29848](#), Ernesto Puerta)
- mgr/dashboard: fix small typos in description message ([pr#30647](#), Tatjana Dehler)
- mgr/dashboard: fix some performance data are not displayed ([issue#39971](#), [pr#28169](#), Kiefer Chang)
- mgr/dashboard: fix sparkline component ([pr#26985](#), Alfonso Martxc3xadnez)
- mgr/dashboard: fix tasks.mgr.dashboard.test\_rgw suite ([pr#33718](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Fix the table mouseenter event handling test ([pr#28879](#), Stephan Mxc3xbcller)
- mgr/dashboard: fix tox test failure ([pr#29125](#), Kiefer Chang)
- mgr/dashboard: Fix translation of variables ([pr#30671](#), Tiago Melo)
- mgr/dashboard: Fix typo in NFS form ([issue#39067](#), [pr#27245](#), Tiago Melo)
- mgr/dashboard: fix visibility of pwdExpirationDate field ([pr#32703](#), Tatjana Dehler)

- mgr/dashboard: Fix zsh support in run-backend-api-tests.sh ([pr#31070](#), Sebastian Wagner)
- mgr/dashboard: Fix npm run fixmod command ([pr#28176](#), Patrick Nawracay)
- mgr/dashboard: Fixes defaultBuilder is not a function ([pr#29420](#), Ricardo Marques)
- mgr/dashboard: Fixes random cephfs tab test failure ([pr#30814](#), Stephan Mxc3xbcller)
- mgr/dashboard: Fixes rbd image purge trash button & modal text ([pr#33321](#), anurag)
- mgr/dashboard: Fixes tooltip behavior ([pr#27153](#), Stephan Mxc3xbcller)
- mgr/dashboard: FixtureHelper ([pr#27157](#), Stephan Mxc3xbcller)
- mgr/dashboard: Form fields do not show error messages/hints ([pr#29043](#), Volker Theile)
- mgr/dashboard: ganesha: Specify the name of the filesystem (create\_path) ([pr#29182](#), David Casier)
- mgr/dashboard: hide daemon table when orchestrator is disabled ([pr#33941](#), Kiefer Chang)
- mgr/dashboard: hide in-use devices when creating OSDs ([pr#31927](#), Kiefer Chang)
- mgr/dashboard: improve device selection modal for creating OSDs ([pr#33081](#), Kiefer Chang)
- mgr/dashboard: Improve hints shown when message.xlf is invalid ([issue#40064](#), [pr#28377](#), Patrick Nawracay)
- mgr/dashboard: Improve NFS Pseudo pattern message ([issue#39327](#), [pr#27653](#), Tiago Melo)
- mgr/dashboard: Improve Notification sidebar ([pr#32895](#), Tiago Melo)
- mgr/dashboard: Improve RestClient error logging ([pr#29794](#), Volker Theile)
- mgr/dashboard: Increase column size on mgr module form ([pr#29107](#), Ricardo Marques)
- mgr/dashboard: install teuthology using pip ([pr#31815](#), Kefu Chai)
- mgr/dashboard: internationalization support with AOT enabled ([pr#30694](#), Tiago Melo, Ricardo Dias)
- mgr/dashboard: Invalid SSO configuration when certificate path does not exist ([pr#31920](#), Ricardo Marques)
- mgr/dashboard: iSCSI GET requests should not be logged ([pr#27813](#), Ricardo

Marques)

- mgr/dashboard: iSCSI targets not available if any gateway is down ([pr#31819](#), Ricardo Marques)
- mgr/dashboard: Isolate each RBD component ([pr#33520](#), Tiago Melo)
- mgr/dashboard: KeyError on dashboard reload ([pr#31469](#), Patrick Seidensal)
- mgr/dashboard: KV-table transforms dates through pipe ([pr#27612](#), Stephan Mxc3xbcller)
- mgr/dashboard: Left align badge datatable columns ([pr#32053](#), Volker Theile)
- mgr/dashboard: list services and daemons ([pr#33531](#), Sage Weil, Kiefer Chang)
- mgr/dashboard: Localization for date picker module ([pr#27275](#), Stephan Mxc3xbcller)
- mgr/dashboard: Make all columns sortable ([pr#27784](#), Stephan Mxc3xbcller)
- mgr/dashboard: make check mypy failure ([pr#33573](#), Volker Theile)
- mgr/dashboard: Make password policy check configurable ([pr#32546](#), Volker Theile)
- mgr/dashboard: Make preventDefault work with 400 errors ([pr#26561](#), Stephan Mxc3xbcller)
- mgr/dashboard: monitoring: improve generic Could not reach external API message ([pr#32648](#), Patrick Seidensal)
- mgr/dashboard: Not able to restrict bucket creation for new user ([pr#33612](#), Volker Theile)
- mgr/dashboard: Optimize portal IPs calculation ([pr#28084](#), Ricardo Marques)
- mgr/dashboard: orchestrator integration initial works ([pr#29127](#), Kiefer Chang)
- mgr/dashboard: OSD custom action button removal ([pr#28095](#), Stephan Mxc3xbcller)
- mgr/dashboard: OSD improvements ([pr#30493](#), Patrick Seidensal)
- mgr/dashboard: pass a list of drive\_group to create\_osds ([pr#33014](#), Kefu Chai)
- mgr/dashboard: Pool form uses different loading spinner ([pr#28649](#), Volker Theile)
- mgr/dashboard: Prevent deletion of iSCSI IQNs with open sessions ([pr#29133](#), Ricardo Marques)
- mgr/dashboard: Prevent KeyError when requesting always\_on\_modules ([pr#30426](#), Volker Theile)
- mgr/dashboard: Process password complexity checks immediately ([pr#32032](#), Volker

Theile, Tatjana Dehler)

- mgr/dashboard: Provide the name of the object being deleted ([pr#30658](#), Ricardo Marques)
- mgr/dashboard: Provide user enable/disable capability ([issue#25229](#), [pr#29046](#), Ricardo Dias, Patrick Nawracay)
- mgr/dashboard: Push Grafana dashboards on startup ([pr#26415](#), Zack Cerza)
- mgr/dashboard: qa: fix RBD test when matching error strings ([pr#29264](#), Ricardo Dias)
- mgr/dashboard: qa: whitelist client eviction warning ([pr#29114](#), Ricardo Dias)
- mgr/dashboard: RBD snapshot name suggestion with local time suffix ([pr#27613](#), Stephan Mxc3xbcller)
- mgr/dashboard: Reduce the number of renders on the tables ([issue#39944](#), [pr#28118](#), Tiago Melo)
- mgr/dashboard: Refactor and cleanup tasks.mgr.dashboard.test\_user ([pr#33743](#), Volker Theile)
- mgr/dashboard: Refactor Python unittests and controller ([pr#31165](#), Volker Theile)
- mgr/dashboard: Reload all CephFS directories ([pr#32552](#), Stephan Mxc3xbcller)
- mgr/dashboard: remove config-opt: read perm. from system roles ([pr#33690](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Remove ngx-store ([pr#33756](#), Tiago Melo)
- mgr/dashboard: remove traceback/version assertions ([pr#31720](#), Ernesto Puerta)
- mgr/dashboard: Remove unused RBD configuration endpoint ([pr#30815](#), Ricardo Marques)
- mgr/dashboard: Remove unused variable ([pr#31785](#), Volker Theile)
- mgr/dashboard: Removes distracting search behavior ([pr#27438](#), Stephan Mxc3xbcller)
- mgr/dashboard: Rename pipe list -> join ([pr#31843](#), Volker Theile)
- mgr/dashboard: Replace IP address validation with Python standard library functions ([pr#26184](#), Ashish Singh)
- mgr/dashboard: Replace ng2-tree with angular-tree-component ([pr#33758](#), Tiago Melo)
- mgr/dashboard: RGW bucket creation when no placement target received ([pr#29280](#),

- alfonsomthd)
- mgr/dashboard: RGW port autodetection does not support Beast RGW frontend ([pr#33060](#), Volker Theile)
- mgr/dashboard: RGW User quota validation is not working correctly ([pr#29132](#), Volker Theile)
- mgr/dashboard: run e2e tests against prod build (jenkins job) ([pr#29198](#), alfonsomthd)
- mgr/dashboard: run-frontend-e2e-tests.sh: allow user defined BASE\_URLxe2x80xa6 ([pr#32211](#), Alfonso Martxc3xadnez)
- mgr/dashboard: select placement target on RGW bucket creation ([pr#28764](#), alfonsomthd)
- mgr/dashboard: Set RO as the default access\_type for RGW NFS exports ([pr#30111](#), Tiago Melo)
- mgr/dashboard: show checkboxes for booleans ([pr#32836](#), Tatjana Dehler)
- mgr/dashboard: show correct RGW user system info ([pr#33206](#), Alfonso Martxc3xadnez)
- mgr/dashboard: Show iSCSI gateways status in the health page ([pr#29112](#), Ricardo Marques)
- mgr/dashboard: smart: smart data read out on down osd causes error popup ([pr#32953](#), Volker Theile)
- mgr/dashboard: Standby Dashboards dont handle all requests properly ([pr#30478](#), Volker Theile)
- mgr/dashboard: Support ceph-iscsi config v9 ([pr#27448](#), Ricardo Marques)
- mgr/dashboard: support multiple DriveGroups when creating OSDs ([pr#32678](#), Kiefer Chang)
- mgr/dashboard: support removing OSDs in OSDs page ([pr#31997](#), Kiefer Chang)
- mgr/dashboard: support setting password hashes ([pr#29138](#), Fabian Bonk)
- mgr/dashboard: tasks: only unblock controller thread after TaskManager thread ([pr#30747](#), Ricardo Dias)
- mgr/dashboard: Throw a more meaningful exception ([pr#32234](#), Volker Theile)
- mgr/dashboard: tox.ini fixes ([pr#30779](#), Alfonso Martxc3xadnez)
- mgr/dashboard: UI fixes ([pr#33171](#), Avan Thakkar)

- mgr/dashboard: Unable to set boolean values to false when default is true ([pr#31738](#), Ricardo Marques)
- mgr/dashboard: unify button/URL actions naming ([issue#37337](#), [pr#26572](#), Ernesto Puerta)
- mgr/dashboard: Unify the look of dashboard charts ([issue#39384](#), [pr#27681](#), Tiago Melo)
- mgr/dashboard: update dashboard CODEOWNERShip ([pr#31193](#), Ernesto Puerta)
- mgr/dashboard: Update tar to v4.4.8 ([pr#28092](#), Kefu Chai)
- mgr/dashboard: update vstart to use new ssl port ([issue#26914](#), [pr#27269](#), Ernesto Puerta)
- mgr/dashboard: Updated octopus image on 404 page ([pr#33920](#), Lenz Grimmer)
- mgr/dashboard: Use booleanText pipe ([pr#26733](#), Volker Theile)
- mgr/dashboard: Use default language when running npm run build ([pr#31563](#), Tiago Melo)
- mgr/dashboard: Use ModalComponent in all modals ([pr#33858](#), Tiago Melo)
- mgr/dashboard: Use Observable in auth.service ([pr#32084](#), Volker Theile)
- mgr/dashboard: Use onCancel on any modal event ([pr#29402](#), Stephan Mxc3xbcller)
- mgr/dashboard: Validate iSCSI controls min/max value ([pr#28942](#), Ricardo Marques)
- mgr/dashboard: Validate iSCSI images features ([pr#27135](#), Ricardo Marques)
- mgr/dashboard: Validate ceph-iscsi config version ([pr#26835](#), Ricardo Marques)
- mgr/dashboard: Various UI issues related to CephFS ([pr#29272](#), Volker Theile)
- mgr/dashboard: Vertically align the Refresh label ([pr#29737](#), Tiago Melo)
- mgr/dashboard: vstart: Fix /dev/tty No such device or address ([pr#31195](#), Volker Theile)
- mgr/dashboard: wait for PG unknown state to be cleared ([pr#33013](#), Tatjana Dehler)
- mgr/dashboard: Watch for pool pgs increase and decrease ([pr#28006](#), Ricardo Dias, Stephan Mxc3xbcller)
- mgr/modules: outsource SSL certificate creation ([pr#33550](#), Patrick Seidensal)
- mgr/orch,cephadm: add timestamps to daemons and services ([pr#33728](#), Sage Weil)
- mgr/orch: add -all-available-devices to orch apply osd ([pr#33990](#), Sage Weil)

- mgr/orch: add missing CLI commands for grafana, alertmanager ([pr#33695](#), Sage Weil)
- mgr/orch: associate addresses with hosts ([pr#33098](#), Sage Weil)
- mgr/orch: ceph orchestrator ... -> ceph orch ... ([pr#33131](#), Sage Weil)
- mgr/orch: ceph upgrade ... -> ceph orch upgrade ... ([pr#34046](#), Sage Weil)
- mgr/orch: collapse SPEC and PLACEMENT columns in orch ls ([pr#33795](#), Sage Weil)
- mgr/orch: dump service spec by name ([pr#33951](#), Michael Fritch)
- mgr/orch: first phase of new cli ([pr#33212](#), Sage Weil)
- mgr/orch: fix host ls ([pr#33486](#), Sage Weil)
- mgr/orch: fix orch ls table spacing ([pr#33586](#), Sage Weil)
- mgr/orch: fix ServiceSpec deserialization error ([pr#33779](#), Kiefer Chang)
- mgr/orch: improve commandline parsing for update\\_\* ([pr#31672](#), Joshua Schmid)
- mgr/orch: include spec ref in ServiceDescription ([pr#33667](#), Sage Weil)
- mgr/orch: make arg hostname, not host ([pr#33474](#), Sage Weil)
- mgr/orch: new cli, phase 2 ([pr#33244](#), Sage Weil)
- mgr/orch: pass unicode string to ipaddress.ip\_network() ([pr#31755](#), Kefu Chai)
- mgr/orch: PlacementSpec: add all\_hosts property ([pr#33465](#), Sage Weil)
- mgr/orch: Properly handle NotImplementedError ([pr#33914](#), Sebastian Wagner)
- mgr/orch: remove ansible and deepsea ([pr#33126](#), Sage Weil)
- mgr/orch: resurrect ServiceDescription, orch ls ([pr#33359](#), Sage Weil)
- mgr/orch: take a single placement argument ([pr#33706](#), Sage Weil)
- mgr/orchestrator,mgr/ssh: add host labels ([pr#31854](#), Sage Weil)
- mgr/orchestrator: Add doc about how to use OrchestratorClientMixin ([pr#32893](#), Sebastian Wagner)
- mgr/orchestrator: Add mypy static type checking ([pr#32010](#), Sebastian Wagner)
- mgr/orchestrator: add optional format param for orchestrator host ls ([pr#31930](#), Kefu Chai)
- mgr/orchestrator: add progress events to all orchestrators ([pr#26654](#), Sebastian Wagner)

- mgr/orchestrator: Add simple scheduler ([pr#32003](#), Joshua Schmid)
- mgr/orchestrator: addr is optional for constructing InventoryNode ([pr#33347](#), Kefu Chai)
- mgr/orchestrator: device lights ([pr#26768](#), Sebastian Wagner, Sage Weil)
- mgr/orchestrator: do not try to iterate through None ([pr#31705](#), Kefu Chai)
- mgr/orchestrator: Document OSD replacement ([pr#29792](#), Sebastian Wagner)
- mgr/orchestrator: fix orch host label rm help text ([pr#33585](#), Sage Weil)
- mgr/orchestrator: Fix raise\_if\_exception for Python 3 ([pr#31015](#), Sebastian Wagner)
- mgr/orchestrator: fix refs property of progresses ([pr#30197](#), Kiefer Chang)
- mgr/orchestrator: fix ceph orch apply -i + yaml cleanup + Completion cleanup ([pr#34001](#), Sebastian Wagner)
- mgr/orchestrator: functools.partial doesnt work for methods ([pr#33432](#), Sebastian Wagner)
- mgr/orchestrator: get\_hosts return HostSpec instead of InventoryDevice ([pr#33258](#), Sebastian Wagner)
- mgr/orchestrator: Make Completions composable ([pr#30262](#), Sebastian Wagner, Tim Serong)
- mgr/orchestrator: make hosts and label args consistent ([pr#32253](#), Sage Weil)
- mgr/orchestrator: Raise more expressive Error, if completion already xe2x80xa6 ([pr#32270](#), Sebastian Wagner)
- mgr/orchestrator: raise\_if\_exception: Add exception type to message ([pr#32574](#), Sebastian Wagner)
- mgr/orchestrator: Remove (add|test|remove)\_stateful\_service\_rule ([pr#26772](#), Sebastian Wagner)
- mgr/orchestrator: set node labels to empty list if none specified ([pr#31914](#), Tim Serong)
- mgr/orchestrator: Split \*\_stateless\_service and add get\_feature\_set ([pr#29063](#), Sebastian Wagner)
- mgr/orchestrator: Substitute hostname for nodename, globally ([pr#33467](#), Sebastian Wagner)
- mgr/orchestrator: unify StatelessServiceSpec and StatefulServiceSpec ([pr#33175](#), Sebastian Wagner)

- mgr/orchestrator: use deepcopy for copying exceptions ([pr#32881](#), Kefu Chai)
- mgr/orchestrator: Use pickle to pass exceptions across sub-interpreters ([pr#33179](#), Sebastian Wagner)
- mgr/orchestrator\_cli: clean up device ls table ([pr#32279](#), Sage Weil)
- mgr/orchestrator\_cli: Fix NFS ([pr#32272](#), Sebastian Wagner)
- mgr/orchestrator\_cli: improve service ls output, sorting ([pr#31539](#), Sage Weil)
- mgr/orchestrator\_cli: set type for orchestrator option ([pr#32189](#), Sage Weil)
- mgr/orchestrator\_cli: sort host list ([pr#33370](#), Sage Weil)
- mgr/orchestrator\_cli: \_update\_mons require host spec only ([pr#32499](#), Sebastian Wagner)
- mgr/progress/module.py: s/events/\_events/ ([pr#29625](#), Kamoltat (Junior) Sirivadhna)
- mgr/rook: Add caching for the Dashboard ([pr#29131](#), Sebastian Wagner, Paul Cuzner)
- mgr/rook: Added missing rgw daemons in service ls ([issue#39171](#), [pr#27491](#), Sebastian Wagner)
- mgr/rook: Added Mypy static type checking ([pr#32127](#), Sebastian Wagner)
- mgr/rook: Fix creation of bluestore OSDs ([issue#39062](#), [pr#27289](#), Sebastian Wagner)
- mgr/rook: Fix error creating OSDs ([pr#33176](#), Juan Miguel Olmo Martxc3xadnez)
- mgr/rook: Fix Python 2 regression ([issue#39250](#), [pr#27516](#), Sebastian Wagner)
- mgr/rook: Fix RGW creation ([issue#39158](#), [pr#27462](#), Sebastian Wagner)
- mgr/rook: misc fixes for orch ps ([pr#33868](#), Sage Weil)
- mgr/rook: provide full path for devices names in inventory ([pr#32654](#), Sage Weil)
- mgr/rook: Remove support for Rook older than v0.9 ([issue#39278](#), [pr#27556](#), Sebastian Wagner)
- mgr/rook: Support other system namespaces ([issue#38799](#), [pr#27290](#), Sebastian Wagner)
- mgr/ssh/tests: fix RGWSpec test ([pr#31983](#), Sage Weil)
- mgr/ssh: add per-service operations: start, stop, restart, redeploy ([pr#31292](#), Sage Weil)
- mgr/ssh: add TemporaryDirectory impl for py2 compat ([pr#31835](#), Sage Weil)

- mgr/ssh: allow passing LV to orchestrator osd create ([pr#31512](#), Sage Weil)
- mgr/ssh: annotate object representation ([pr#31602](#), Joshua Schmid)
- mgr/ssh: cache service inventory ([pr#31385](#), Sage Weil)
- mgr/ssh: deploy and remove rgw daemons ([pr#31303](#), Sage Weil)
- mgr/ssh: deploy rbd-mirror daemons ([pr#31493](#), Sage Weil)
- mgr/ssh: fix redeploy ([pr#31613](#), Sage Weil)
- mgr/ssh: fix service\_action, remove\_osds ([pr#31952](#), Sage Weil)
- mgr/ssh: Fix various Python issues ([pr#31524](#), Volker Theile)
- mgr/ssh: Ignore ssh-config file ([pr#31710](#), Volker Theile)
- mgr/ssh: implement blink\_device\_light ([pr#31438](#), Sage Weil)
- mgr/ssh: implement service ls ([pr#31169](#), Sage Weil)
- mgr/ssh: improve service ls ([pr#31828](#), Sage Weil)
- mgr/ssh: Install SSH public key in Vagrantfile box fails ([pr#31519](#), Volker Theile)
- mgr/ssh: optionally specify service names ([pr#31537](#), Sage Weil)
- mgr/ssh: packaged-ceph-daemon mode; ssh key mgmt ([pr#31698](#), Sage Weil)
- mgr/ssh: Port raising exceptions from completion handlers to Py2 ([pr#31940](#), Sebastian Wagner)
- mgr/ssh: raise RuntimeError when ceph-daemon invocation fails ([pr#31420](#), Sage Weil)
- mgr/ssh: remove superfluous parameters ([pr#31462](#), Joshua Schmid)
- mgr/ssh: set up dummy known\_hosts file ([pr#31721](#), Sage Weil)
- mgr/ssh: take IP, CIDR, or addrvec for new mon(s) ([pr#31505](#), Sage Weil)
- mgr/ssh: upgrade check command ([pr#31827](#), Sage Weil)
- mgr/ssh: test\_mon\_update needs to set a mon name ([pr#31933](#), Sebastian Wagner)
- mgr/telemetry: anonymizing smartctl report itself ([pr#33029](#), Yaarit Hatuka)
- mgr/telemetry: dict.pop() errs on nonexistent key ([pr#30854](#), Dan Mick)
- mgr/telemetry: fix log typo ([pr#31984](#), Sage Weil)
- mgr/test\_orchestrator: Allow initializing dummy data ([pr#29595](#), Kiefer Chang)

- mgr/test\_orchestrator: fix tests ([pr#33541](#), Sage Weil)
- mgr/test\_orchestrator: Fix TestWriteCompletion object has no attribute id ([pr#27607](#), Sebastian Wagner)
- mgr/test\_orchestrator: fix update\_mgrs assert ([pr#32417](#), Sage Weil)
- mgr/volumes: add arg to fs volume create for mds daemons placement ([pr#33441](#), Daniel-Pivonka)
- mgr: Add get\_rates\_from\_data to mgr\_util.py ([pr#28603](#), Stephan Mxc3xbcller)
- mgr: add rbd profiles to support rbd\_support module commands ([pr#30912](#), Jason Dillaman)
- mgr: better error handling when reading option ([pr#32730](#), Kefu Chai)
- mgr: ceph fs status support json format ([pr#30985](#), Erqi Chen)
- mgr: change perf-counter precision to float ([pr#30400](#), Ernesto Puerta)
- mgr: check for unicode passed to set\_health\_checks() ([pr#29117](#), Kefu Chai)
- mgr: cleanup idle debug log at level 4 ([pr#29164](#), Sebastian Wagner)
- mgr: close restful socket after exec ([pr#32396](#), liushi)
- mgr: Configure Py root logger for Mgr modules ([pr#27069](#), Volker Theile)
- mgr: do not reset reported if a new metric is not collected ([pr#30285](#), Ilsoo Byun)
- mgr: drop session with Ceph daemon when not ready ([pr#31899](#), Patrick Donnelly)
- mgr: fix a few bugs with teh pgp\_num adjustments ([pr#27875](#), Sage Weil)
- mgr: fix ceph native option value types ([pr#29855](#), Sage Weil)
- mgr: fix debug typo ([pr#31900](#), Patrick Donnelly)
- mgr: fix errors on using a reference in a Lambda function ([pr#31786](#), Willem Jan Withagen)
- mgr: fix reporting of per-module logging options to mon ([pr#33897](#), Sage Weil)
- mgr: fix weird health-alert daemon key ([pr#30617](#), xie xingguo)
- mgr: handle race with finisher after shutdown ([pr#31620](#), Patrick Donnelly)
- mgr: Improve internal python to c++ interface ([pr#32554](#), David Zafman)
- mgr: install tox deps from wheelhouse ([pr#30034](#), Kefu Chai)

- mgr: mgr, osd: osd df by pool ([pr#28629](#), xie xingguo)
- mgr: mgr/ActivePyModules: behave if a module queries a devid that does not exist ([pr#31291](#), Sage Weil)
- mgr: mgr/ActivePyModules: drop GIL while we wait for mon reply in set\_store, set\_config ([issue#39335](#), [pr#27619](#), Sage Weil)
- mgr: mgr/ActivePyModules: handle\_command - fix broken lock ([issue#39235](#), [pr#27485](#), xie xingguo)
- mgr: mgr/balancer: avoid pulling pg\_dump twice ([pr#32266](#), xie xingguo)
- mgr: mgr/balancer: eliminate usage of MS infrastructure for upmap mode ([pr#32289](#), xie xingguo)
- mgr: mgr/balancer: enable pg\_upmap cli for future use ([pr#30560](#), xie xingguo)
- mgr: mgr/balancer: fix fudge ([pr#27994](#), xie xingguo)
- mgr: mgr/balancer: fix initial weight-set value for newly created osds ([pr#28251](#), xie xingguo)
- mgr: mgr/balancer: Python 3 compatibility fix ([issue#38831](#), [pr#27076](#), Marius Schiffer)
- mgr: mgr/balancer: python3 compatibility issue ([pr#30987](#), Mykola Golub)
- mgr: mgr/balancer: upmap\_max\_iterations -> upmap\_max\_optimizations; behave as it is per pool ([pr#30591](#), xie xingguo)
- mgr: mgr/BaseMgrModule: tolerate Int or Long for health count ([pr#29806](#), Sage Weil)
- mgr: mgr/BaseMgrModule: use PyInt\_Check() to compatible with py2 ([pr#29831](#), Kefu Chai)
- mgr: mgr/BaseMgrStandbyModule: drop GIL in ceph\_get\_module\_option() ([pr#30625](#), Kefu Chai)
- mgr: mgr/cephadm: custom certificates for Grafana deployment ([pr#33614](#), Patrick Seidensal)
- mgr: mgr/cephadm: support (point release) upgrades ([pr#32006](#), Sage Weil)
- mgr: mgr/crash: Calculate and add stack\_sig to metadata ([pr#31394](#), Dan Mick)
- mgr: mgr/crash: fix crash ls[-new] sorting ([pr#31973](#), Sage Weil)
- mgr: mgr/DaemonServer: handle caps more carefully ([pr#26903](#), xie xingguo)
- mgr: mgr/DaemonServer: handle\_conf\_change - fix broken locking ([issue#38899](#),

[pr#27184](#), xie xingguo)

- mgr: mgr/DaemonServer: refactor pgp\_num changes throttling ([pr#27891](#), Kefu Chai)
- mgr: mgr/DaemonServer: safe-to-destroy - do not consider irrelevant pgs ([pr#27962](#), xie xingguo)
- mgr: mgr/DaemonServer: skip adjusting pgp\_num when merging is in-progress ([pr#30139](#), xie xingguo)
- mgr: mgr/dashboard: Do not default to admin as Admin Resource ([issue#39338](#), [pr#27626](#), Wido den Hollander)
- mgr: mgr/dashboard: Handle always-on Ceph Manager modules correctly ([pr#30142](#), Volker Theile)
- mgr: mgr/dashboard: integrate progress mgr module events into dashboard tasks list ([pr#29048](#), Ricardo Dias)
- mgr: mgr/dashboard: Manager should complain about wrong dashboard certificate ([pr#27036](#), Volker Theile)
- mgr: mgr/deepsea: return ganesha and iscsi endpoint URLs ([pr#27336](#), Tim Serong)
- mgr: mgr/deepsea: use ceph\_volume output in get\_inventory() ([pr#26966](#), Tim Serong)
- mgr: mgr/devicehealth: ensure we dont store empty objects ([pr#31474](#), Sage Weil)
- mgr: mgr/devicehealth: Fix python 3 incompatibility ([issue#38939](#), [pr#27172](#), Marius Schiffer)
- mgr: mgr/devicehealth: set default monitoring to on ([pr#33091](#), Sage Weil, Yaarit Hatuka)
- mgr: mgr/diskprediction: Add diskprediction local plugin dependencies ([pr#25530](#), Rick Chen)
- mgr: mgr/diskprediction\_cloud: Correct base64 encode translate table ([issue#38848](#), [pr#27113](#), Rick Chen)
- mgr: mgr/diskprediction\_cloud: refactor timeout() decorator ([pr#31176](#), Kefu Chai)
- mgr: mgr/hello: some clean up and modernization ([pr#29514](#), Sage Weil)
- mgr: mgr/influx: try to call close() ([issue#40174](#), [pr#28427](#), Kefu Chai)
- mgr: mgr/insights: fix prune-health-history ([pr#32973](#), Sage Weil)
- mgr: mgr/k8sevents: Add mgr module for kubernetes event integration ([pr#29520](#), Paul Cuzner)

- mgr: mgr/k8sevents: Add support for remote kubernetes ([pr#30482](#), Paul Cuzner)
- mgr: mgr/Mgr: kill redundant sub\_unwant call ([pr#26950](#), xie xingguo)
- mgr: mgr/MgrMonitor: print pending.always\_on\_modules before updating it ([pr#29917](#), Kefu Chai)
- mgr: mgr/orch: logging - handle lists output ([pr#32879](#), Shyukri Shyukriev)
- mgr: mgr/orchestrator: Add cache for Inventory and Services ([pr#28213](#), Tim Serong, Sebastian Wagner)
- mgr: mgr/orchestrator\_cli: pass default value to req=False params ([pr#31314](#), Kefu Chai)
- mgr: mgr/osd\_support: new module for osd utility ([pr#32677](#), Joshua Schmid)
- mgr: mgr/pg\_autoscaler: calculate pool\_pg\_target using pool size ([pr#32592](#), Dan van der Ster)
- mgr: mgr/pg\_autoscaler: fix pool\_logical\_used ([pr#29986](#), Ansgar Jazdzewski)
- mgr: mgr/pg\_autoscaler: Fix python3 incompatibility ([issue#38626](#), [pr#27079](#), Marius Schiffer)
- mgr: mgr/pg\_autoscaler: fix race with pool deletion ([pr#29807](#), Sage Weil)
- mgr: mgr/pg\_autoscaler: treat target ratios as weights ([pr#33035](#), Josh Durgin)
- mgr: mgr/progress & mgr/pg\_autoscaler: Added Pg Autoscaler Event ([pr#29035](#), Kamoltat (Junior) Sirivadhna)
- mgr: mgr/progress: Add integration to pybind/mgr/tox.ini ([pr#32985](#), Sebastian Wagner)
- mgr: mgr/progress: Add recovery event when OSD marked in ([pr#28498](#), Kamoltat (Junior) Sirivadhna)
- mgr: mgr/progress: added the time an event has been in progress ([pr#28907](#), Kamoltat (Junior) Sirivadhna)
- mgr: mgr/progress: Bug fix complete event when OSD marked in ([pr#28695](#), Kamoltat (Junior) Sirivadhna)
- mgr: mgr/progress: clamp pg recovery ratio to 0 ([pr#29126](#), xie xingguo)
- mgr: mgr/progress: estimated remaining time for events ([pr#30615](#), xie xingguo)
- mgr: mgr/progress: Look at PG state when PG epoch >= OSDMap epoch ([pr#28368](#), Kamoltat (Junior) Sirivadhna)
- mgr: mgr/progress: remove since from duration string ([pr#31007](#), Kefu Chai)

- mgr: mgr/prometheus: Add mgr metadata to prometheus exporter module ([pr#28372](#), Paul Cuzner)
- mgr: mgr/prometheus: assign a value to osd\_dev\_node when obj\_store is not filestore or bluestore ([pr#30534](#), jiahuizeng)
- mgr: mgr/prometheus: Cast collect\_timeout (scrape\_interval) to float ([pr#29382](#), Ben Meekhof)
- mgr: mgr/prometheus: Fix KeyError in get\_mgr\_status ([pr#30421](#), Sebastian Wagner)
- mgr: mgr/prometheus: replace whitespaces in metrics names ([pr#27722](#), Alfonso Martxc3xadnez)
- mgr: mgr/PyModule: correctly remove config options ([pr#31807](#), Tim Serong)
- mgr: mgr/PyModuleRegistry: log error if we cant find any modules to load ([pr#28055](#), Tim Serong)
- mgr: mgr/restful: allow shutdown before weve fully started up ([pr#32004](#), Sage Weil)
- mgr: mgr/restful: do not use filter() for list ([pr#27925](#), Kefu Chai)
- mgr: mgr/restful: jsonify lists instead of maps ([pr#32421](#), Kefu Chai)
- mgr: mgr/restful: requests api adds support multiple commands ([pr#31152](#), Duncan Chiang)
- mgr: mgr/status: fix ceph osd status ZeroDivisionError ([pr#28797](#), simon gao)
- mgr: mgr/telemetry: add last\_upload to status ([pr#33125](#), Yaarit Hatuka)
- mgr: mgr/telemetry: change crash dict to a list ([pr#27631](#), Dan Mick)
- mgr: mgr/telemetry: channels ([pr#28847](#), Sage Weil)
- mgr: mgr/telemetry: check get\_metadata return val ([pr#33051](#), Yaarit Hatuka)
- mgr: mgr/telemetry: clear the event after being awaken by it ([pr#29546](#), Kefu Chai)
- mgr: mgr/telemetry: exclude hostname field in crash reports ([pr#27693](#), Sage Weil)
- mgr: mgr/telemetry: fix and document proxy usage ([pr#33575](#), Lars Marowsky-Bree)
- mgr: mgr/telemetry: fix device serial number anonymization ([pr#32492](#), Yaarit Hatuka)
- mgr: mgr/telemetry: include any config options that are customized ([pr#29334](#), Sage Weil)

- mgr: mgr/telemetry: include device health telemetry ([pr#30724](#), Sage Weil)
- mgr: mgr/telemetry: re-opt-in when telemetry content changes; nag on major releases ([pr#29337](#), Sage Weil)
- mgr: mgr/telemetry: salt osd ids too ([pr#29358](#), Sage Weil)
- mgr: mgr/telemetry: specify license when opting in ([pr#29340](#), Sage Weil)
- mgr: mgr/volumes: do not import unused module ([pr#28875](#), Kefu Chai)
- mgr: mgr/zabbix Added pools discovery and per-pool statistics ([pr#26152](#), Dmitriy Rabotjagov)
- mgr: mgr/zabbix: Adds possibility to send data to multiple zabbix servers ([issue#38409](#), [pr#26547](#), slivik, Jakub Sliva)
- mgr: mgr/zabbix: encode string for Python 3 compatibility ([pr#28624](#), Nathan Cutler)
- mgr: mgr/zabbix: Fix raw\_bytes\_used key name ([pr#28058](#), Dmitriy Rabotjagov)
- mgr: mgr/zabbix: Fix typo in key name for PGs in backfill\_wait state ([issue#39666](#), [pr#28057](#), Wido den Hollander)
- mgr: missing lock release in DaemonServer::handle\_report() ([issue#42169](#), [pr#30706](#), Venky Shankar)
- mgr: module logging infrastructure ([pr#30961](#), Ricardo Dias)
- mgr: more GIL fixes ([issue#39040](#), [pr#27280](#), xie xingguo)
- mgr: pybind/mgr/balancer/module.py: add max/min info in stats\_by\_root ([pr#30432](#), Yang Honggang)
- mgr: pybind/mgr/pg\_autoscaler: implement shutdown method ([pr#31398](#), Patrick Donnelly)
- mgr: pybind/mgr/restful: use dict.items() for py3 compatible ([pr#29356](#), Kefu Chai)
- mgr: pybind/mgr: Cancel output color control ([pr#31427](#), Zheng Yin)
- mgr: pybind/mgr: convert str to int using int() ([pr#27926](#), Kefu Chai)
- mgr: pybind/mgr: Make it easier to create a Module instance without the mgr ([pr#31969](#), Sebastian Wagner)
- mgr: pybind/mgr: Remove code duplication ([issue#40698](#), [pr#28986](#), Sebastian Wagner)
- mgr: pybind/mgr: add mgr\_module.py and mgr\_util.py to mypy ([pr#32597](#), Sebastian Wagner)

Wagner)

- mgr: Python cleanup and type check ([pr#31559](#), Volker Theile)
- mgr: qa/mgr/progress: fix timeout error when waiting for osd in event ([pr#30095](#), Ricardo Dias)
- mgr: re-enable mds scrub status info in ceph status ([issue#42835](#), [pr#32657](#), Venky Shankar)
- mgr: Reduce logging noise when handling commands ([pr#29305](#), Sebastian Wagner)
- mgr: Release GIL before calling OSDMap::calc\_pg\_upmaps() ([pr#31064](#), David Zafman)
- mgr: remove unused variable pool\_name ([pr#28340](#), Alex Wu)
- mgr: restful: Expose perf counters ([pr#27885](#), Boris Ranto)
- mgr: restful: Query nodes\_by\_id for items ([pr#31153](#), Boris Ranto)
- mgr: return perf\_counters data timestamps in nanosecs ([pr#28882](#), Ricardo Dias)
- mgr: Revert mgr/DaemonServer: safe-to-destroy - do not consider irrelevant pgs ([pr#32203](#), xie xingguo)
- mgr: set hostname in DeviceState::set\_metadata() ([pr#30448](#), Kefu Chai)
- mgr: simply exit on SIGINT or SIGTERM ([pr#32051](#), Sage Weil)
- mgr: telemetry/server: misc fixes ([pr#29365](#), user.email, Sage Weil)
- mgr: telemetry: misc scripts ([pr#29781](#), sage@newdream.net, Sage Weil)
- mgr: template metrics collection interface ([pr#29214](#), Venky Shankar)
- mgr: update hostname when we already have the daemon state from the same entity ([pr#33752](#), Kefu Chai)
- mgr: use a struct for DaemonKey ([pr#30635](#), Kefu Chai)
- mgr: use ipv4 default when ipv6 was disabled ([pr#28246](#), kungf)
- mgr: use new MMgrCommand for CLI commands sent to mgr ([pr#30155](#), Sage Weil)
- mgr: zabbix triggers never triggered due to wrong trigger function ([pr#26146](#), Sebastiaan Nijhuis)
- mgr: \_exit(0) from signal handler even if we are standby ([pr#31685](#), Sage Weil)
- mon,rbd,tests: mon,test: silence warnings from GCC and test ([pr#28250](#), Kefu Chai)
- mon,tests: qa/tasks: Fix ambiguous store\_thrash, thrash\_store ([issue#39159](#), [pr#27542](#), Jos Collin)

- mon/tools: monmaptool: added -addv option to usage description ([pr#29307](#), Ricardo Dias)
- mon/MonClient: fix mon tell to older mons ([pr#31121](#), Sage Weil)
- mon/OSDMonitor.cc: Allow pool set target\_max\\_(objects/bytes) with SI/IEC units ([pr#31010](#), Prashant D)
- mon/OSDMonitor: osd add-no{up,down,in,out} - remove state checker ([pr#27605](#), xie xingguo)
- mon/pgmap: fix bluestore alerts output ([pr#30342](#), Igor Fedotov)
- mon: add ability to mute health alerts ([pr#29422](#), Sage Weil)
- mon: add mon, osd, mds ok-to-stop and related commands ([pr#27146](#), Sage Weil)
- mon: add ceph osd info to obtain info on osds rather than parsing osd dump ([pr#26724](#), Joao Eduardo Luis)
- mon: allow running without a config file ([pr#30498](#), Joao Eduardo Luis)
- mon: always enable pg\_autoscaler ([pr#29072](#), Sage Weil)
- mon: disable min pg per osd warning ([pr#30352](#), Sage Weil)
- mon: Dont put session during feature change ([pr#32365](#), Brad Hubbard)
- mon: dump json from sessions asok/tell command ([pr#32974](#), Sage Weil)
- mon: elector: return after triggering a new election ([pr#32981](#), Greg Farnum)
- mon: ensure prepare\_failure() marks no\_reply on op ([pr#28177](#), Joao Eduardo Luis)
- mon: fix INCOMPAT\_OCTOPUS feature number ([pr#27622](#), Sage Weil)
- mon: fix misc asok commands ([pr#30859](#), Sage Weil, Patrick Donnelly)
- mon: fix off-by-one rendering progress bar ([pr#28268](#), Sage Weil)
- mon: fix tell command description (and ceph CLI help behavior) ([pr#33135](#), Sage Weil)
- mon: fix tell to hybrid octopus/pre-octopus mons ([pr#31138](#), Sage Weil)
- mon: fix/improve mon sync over small keys ([pr#31581](#), Sage Weil)
- mon: Get session\_map\_lock before remove\_session ([pr#33682](#), Xiaofei Cui)
- mon: Improve health status for backfill\_toofull and recovery\_toofull ([pr#28204](#), David Zafman)
- mon: Improvements to slow heartbeat health messages ([pr#32342](#), David Zafman)

- mon: make ceph -s much more concise ([pr#29493](#), Sage Weil)
- mon: make compact tell command, and add deprecate/obsolete check for tell commands ([pr#31722](#), Kefu Chai)
- mon: make mon\_osd\_down\_out\_subtree\_limit update at runtime ([pr#27517](#), Sage Weil)
- mon: mon/ConfigMonitor: make config reset idempotent ([pr#27155](#), xie xingguo)
- mon: mon/ConfigMonitor: make num of config reset optional; allow target version 0 ([pr#27090](#), xie xingguo)
- mon: mon/HealthMonitor: remove unused label ([pr#29749](#), Kefu Chai)
- mon: mon/MonClient: weight-based mon selection ([pr#26940](#), xie xingguo)
- mon: mon/Monitor: no need to create a local variable for capturing it ([pr#28744](#), Kefu Chai)
- mon: mon/MonMap: always set mon priority; add it to dump ([pr#26975](#), xie xingguo)
- mon: mon/OSDMonitor: crush node flags - two fixes; add tests ([pr#27719](#), xie xingguo)
- mon: mon/OSDMonitor: fix off-by-one when updating new\_last\_in\_change ([pr#28568](#), xie xingguo)
- mon: mon/OSDMonitor: report pg[pgp]\_num\_target instead of pg[pgp]\_num ([issue#40193](#), [pr#28490](#), xie xingguo)
- mon: mon/OSDMonitor: trim not-longer-exist failure reporters ([pr#30200](#), NancySu05)
- mon: mon/OSDMonitor: use initializer\_list<> for {si, iec}\_options ([pr#31175](#), Kefu Chai)
- mon: mon/PGMap: fix incorrect pg\_pool\_sum when delete pool ([pr#31560](#), luo rixin)
- mon: optionally bind to public\_addrv (instead of public\_addr or public\_network) ([pr#31501](#), Sage Weil)
- mon: paxos: empty pending\_finishers before retrying any of committingxe2x80xa6 ([issue#39484](#), [pr#27877](#), Greg Farnum)
- mon: print FSSMap regardless of file system count ([pr#32307](#), Patrick Donnelly)
- mon: quiet devname noise ([pr#27313](#), Sage Weil)
- mon: remove the restriction of address type in init\_with\_hosts ([pr#31691](#), Hao Xiong)
- mon: Revert mon/OSDMonitor: report pg[pgp]\_num\_target instead of

pg[pgp]\_xe2x80xa6 (pr#28567, xie xingguo)

- mon: set recovery\_priority, pg\_num\_min, pg\_autoscale\_bias via fs new command (pr#29180, Sage Weil)
- mon: should not take non-tell commands as tell ones (pr#32517, Kefu Chai)
- mon: show no[deep-]scrub flags per pool in the status (issue#38029, pr#26488, Mohamad Gebai)
- mon: show pool id in pool ls command (issue#40287, pr#28488, Chang Liu)
- mon: Split Elector into message-passing and logic/state components (pr#28727, Greg Farnum)
- mon: stash newer map on bootstrap when addr doesnt match (pr#33418, Sage Weil)
- mon: take the mon lock in handle\_conf\_change (issue#39625, pr#28018, huangjun)
- mon: use non-obsolete mon scrub cmd (pr#32510, Patrick Donnelly)
- mon:C\_AckMarkedDown has not handled the Callback Arguments (pr#29624, NancySu05)
- monitoring: fix prometheus alert for full pools (pr#32325, Thomas Kriechbaumer)
- monitoring: fix RGW grafana chart Average GET/PUT Latencies (pr#33839, Alfonso Martxc3xadnez)
- monitoring: restore lost fix for pool full alert (pr#33655, Patrick Seidensal)
- monitoring: SNMP OID per every Prometheus alert rule (pr#27978, Volker Theile)
- monitoring: wait before firing osd full alert (pr#31711, Patrick Seidensal)
- msg/async, v2: make the reset\_recv\_state() unconditional (issue#40115, pr#28453, Sage Weil, Radoslaw Zarzynski)
- msg/async/AsyncConnection: optimize check loopback connection (pr#26923, Jianpeng Ma)
- msg/async/dpdk: destroy fd in do\_request (pr#32690, Chunsong Feng, luo rixin)
- msg/async/dpdk: Fix build when DPDK enabled (pr#33203, Jun Su)
- msg/async/dpdk: fix compilation errors when WITH\_DPDK=on (pr#31840, Chunsong Feng)
- msg/async/dpdk: fix complie errors from fix FTBFS (pr#30086, yehu)
- msg/async/dpdk: fix FTBFS (pr#28763, Kefu Chai)
- msg/async/dpdk: Fix infinite loop when sending packets (pr#32691, Chunsong Feng, luo rixin)

- msg/async/dpdk: fix SEGV caused by zero length packet ([pr#31876](#), Chunsong Feng)
- msg/async/dpdk: Fix the overflow while parsing dpdk coremask ([pr#32173](#), Hu Ye, Chunsong Feng, luo rixin)
- msg/async/DPDK: refactor set\_rss\_table to support DPDK 19.05 ([pr#32170](#), Chunsong Feng, luo rixin)
- msg/async/EventEpoll: set EPOLLET flag on del\_event() ([pr#26926](#), Roman Penyaev)
- msg/async/ProtocolV1: avoid unnecessary bufferlist::swap ([pr#30125](#), Jianpeng Ma)
- msg/async/ProtocolV2: make v2 work on rdma ([pr#27022](#), Jianpeng Ma)
- msg/async/ProtocolV2: optimize check state by replace ([pr#26812](#), Jianpeng Ma)
- msg/async/rdma: add an option for choosing different RoCE protocol ([pr#31517](#), Changcheng Liu)
- msg/async/rdma: do not init mutex before lockdeps is ready ([pr#31532](#), Kefu Chai)
- msg/async/rdma: fix memory leak ([pr#27574](#), Changcheng Liu)
- msg/async/rdma: set/get silence warning ([pr#26581](#), Kefu Chai)
- msg/async/rdma: unblock event center if the peer is down when connecting ([pr#31109](#), Peng Liu)
- msg/async: add comments for commit 294c41f18adada6a ([pr#28667](#), Jianpeng Ma)
- msg/async: add timeout for connections which are not ready ([issue#38493](#), [issue#37499](#), [pr#27337](#), xie xingguo)
- msg/async: avoid creating unnecessary AsyncConnectionRef ([pr#27323](#), Patrick Donnelly)
- msg/async: Dont dec(msgr\_active\_connections) if conn still in acceptxe2x80xa6 ([pr#29836](#), Jianpeng Ma)
- msg/async: Dont miss record l\_msgr\_running\_recv\_time if pendingReadxe2x80xa6 ([pr#27734](#), Jianpeng Ma)
- msg/async: drop zero\_copy\_read() & co from ConnectedSocket ([pr#28921](#), Radoslaw Zarzynski)
- msg/async: fix typo in Errormessage ([pr#31825](#), Willem Jan Withagen)
- msg/async: mark down local\_connection before draining the stack ([pr#32732](#), Radoslaw Zarzynski)
- msg/async: move submit\_message() into send\_to() ([pr#30883](#), Jianpeng Ma)

- msg/async: narrow scope of AsyncMessenger::lock in fun connect\_to ([pr#30840](#), Jianpeng Ma)
- msg/async: No need lock for func \_filter\_addrs ([pr#31995](#), Jianpeng Ma)
- msg/async: no-need set connection for Message ([pr#27766](#), Jianpeng Ma)
- msg/async: open() should be called with connection locked ([pr#33015](#), Roman Penyaev)
- msg/async: perform recv reset immediately if called inside EC ([pr#33742](#), Radoslaw Zarzynski)
- msg/async: remove unused code ([pr#30833](#), Jianpeng Ma)
- msg/async: rename outcoming\_bl -> outgoing\_bl in AsyncConnection ([pr#30709](#), Radoslaw Zarzynski)
- msg/async: reset the V1s session\_security in proper EventCenter ([pr#32352](#), Radoslaw Zarzynski)
- msg/async: resolve gcc warning ([pr#27414](#), Patrick Donnelly)
- msg/async: skip repeat calc crc header in Message::encode ([pr#26534](#), Jianpeng Ma)
- msg/async: update refcount and perf counter properly ([pr#31929](#), Jianpeng Ma)
- msg/async: use faster clear method to delete containers ([pr#27324](#), Patrick Donnelly)
- msg/Message: Remove used code about XioMessenger ([pr#28719](#), Jianpeng Ma)
- msg: add func is\_blackhole to reduce duplicated code ([pr#30356](#), Jianpeng Ma)
- msg: add some anonymous connection infrastructure ([pr#30223](#), Sage Weil)
- msg: default to debug\_ms=0 ([pr#26936](#), Sage Weil)
- msg: fix addr2 encoding for sockaddrs ([issue#40114](#), [pr#28379](#), Jeff Layton)
- msg: fix comments in Messenger.h after the set -> std::set switch ([pr#30693](#), Radoslaw Zarzynski)
- msg: output peer address when detecting bad CRCs ([issue#39367](#), [pr#27658](#), Greg Farnum)
- msg: remove unused header file in Messenger.h ([pr#27086](#), Jianpeng Ma)
- msg: remove xiomessenger ([pr#27021](#), Sage Weil)
- msg: set\_require\_authorizer on messenger, not dispatcher ([pr#27832](#), Sage Weil)
- orchestrator: usability fixes ([pr#33118](#), Yehuda Sadeh)

- os/bluestore, comon, erasure-code: chmod -x source files ([pr#31179](#), Sage Weil)
- os/bluestore: default bluestore\_block\_size 1T -> 100G ([pr#32043](#), Sage Weil)
- os/kstore: do not cache in-fight stripes on read ops to avoid leaks ([issue#39665](#), [pr#32538](#), Chang Liu)
- os/memstore, crimson/os: introduce memstore\_debug OMIT\_BLOCK\_DEVICE\_WRITE ([pr#28601](#), Radoslaw Zarzynski)
- osd: a few fixes for the removed\_snaps changes ([pr#28865](#), Sage Weil)
- osd: accident of rollforward may need to mark pglog dirty ([issue#40403](#), [pr#28621](#), Zengran Zhang)
- osd: add a copy-from2 operation that includes truncate\\_{seq,size} parameters ([pr#31728](#), Luis Henriques)
- osd: add ceph osd stop <osd.nnn> command ([pr#27595](#), xie xingguo)
- osd: add cls\_cxx\_map\_remove\_range() ([issue#19975](#), [pr#15183](#), Casey Bodley)
- osd: add common smartctl output to JSON output ([pr#30408](#), Patrick Seidensal)
- osd: add device\_id to list\_devices to help get smart info easily ([pr#29548](#), Song Shun)
- osd: add duration field to dump\_historic\_ops method ([pr#28801](#), Deepika Upadhyay)
- osd: add flag to prevent truncate\_seq copy in copy-from operation ([pr#25374](#), Luis Henriques)
- osd: add hdd and ssd variants for osd\_recovery\_max\_active ([pr#28677](#), Sage Weil)
- osd: add log information to record the cause of do\_osd\_ops failure ([issue#41210](#), [pr#29787](#), NancySu05)
- osd: add osd\_fast\_shutdown option (default true) ([pr#31677](#), Sage Weil)
- osd: Again remove deprecated full/nearfull from osdmap ([pr#32506](#), David Zafman)
- osd: Allow 64-char hostname to be added as the host in CRUSH ([pr#32947](#), Michal Skalski)
- osd: allow EC PGs to do recovery below min\_size ([issue#18749](#), [pr#17619](#), Chang Liu, Greg Farnum)
- osd: allow rados write ops to return data and error codes ([pr#30581](#), Sage Weil)
- osd: always initialize local variable ([pr#29757](#), Kefu Chai)
- osd: assert that write ops have result==0 and no payload ([pr#30191](#), Sage Weil)

- osd: automatically repair replicated replica on pulling error ([issue#39101](#), [pr#26806](#), xie xingguo, David Zafman)
- osd: avoid prep\_object\_replica\_pushes() on clone object when head missing ([issue#39286](#), [pr#27575](#), Zengran Zhang)
- osd: Better error message when OSD count is less than osd\_pool\_default\_size ([issue#38617](#), [pr#27806](#), Sage Weil, zjh)
- osd: Change osd op queue cut off default to high ([pr#30441](#), Anthony DAtri)
- osd: clean up osdmap sharing ([pr#27932](#), Sage Weil)
- osd: clear osd op reply output only when writes success ([issue#38492](#), [pr#26652](#), huangjun)
- osd: clear PG\_STATE\_CLEAN when repair object ([pr#29756](#), Zengran Zhang)
- osd: copy (dont move) pg list when sending beacon ([issue#40377](#), [pr#28566](#), Sage Weil)
- osd: copy ObjectOperation::BufferUpdate::Write::fadvise\_flag to ceph::os::Transaction ([pr#29944](#), Xuehan Xu)
- osd: copyfrom omitted to set mtime ([pr#28581](#), Zengran Zhang)
- osd: correct a local variable type ([pr#26672](#), Kefu Chai)
- osd: Diagnostic logging for upmap cleaning ([pr#32663](#), David Zafman)
- osd: dispatch peering messages as messages, inside the PG lock ([pr#29820](#), Sage Weil)
- osd: dispatch\_context and queue split finish on early bail-out ([pr#32942](#), Sage Weil)
- osd: do not hold osd\_lock while requeuing snaps to purge ([pr#28941](#), Sage Weil)
- osd: do not invalidate clear\_regions of missing item at boot ([pr#29755](#), xie xingguo)
- osd: dont carry PGLSFilter between multiple ops in MOSDOP ([pr#29575](#), Radoslaw Zarzynski)
- osd: Dont evict after a flush if intersecting scrub range ([issue#38840](#), [pr#27209](#), David Zafman)
- osd: Dont include user changeable flag in snaptrim related assert ([issue#38124](#), [pr#27830](#), David Zafman)
- osd: Dont randomize deep scrubs when noscrub set ([issue#40198](#), [pr#28443](#), David Zafman)

- osd: drop unnecessary includes of messages/MOSDPGTrim.h ([pr#33660](#), Radoslaw Zarzynski)
- osd: Fix assert in the case that snapset is missing ([pr#29941](#), David Zafman)
- osd: fix possible crash on sending dynamic perf stats report ([pr#30454](#), Mykola Golub)
- osd: fix racy accesses to OSD::osdmap ([pr#33336](#), Radoslaw Zarzynski)
- osd: fix the missing default value m=2 of reed\_sol\_r6\_op in profile ([pr#29892](#), Yan Jun)
- osd: Fix the way that auto repair triggers after regular scru ([issue#40073](#), [issue#40530](#), [pr#28334](#), David Zafman)
- osd: fix wrong arguments when dropping refcount ([pr#29348](#), Myoungwon Oh)
- osd: Give recovery for inactive PGs a higher priority ([issue#38195](#), [pr#27503](#), David Zafman)
- osd: give recovery ops initialized by client op a higher priority ([pr#28418](#), xie xingguo)
- osd: implement per-pg leases to avoid stale reads ([pr#29236](#), Sage Weil)
- osd: Improve dump\_pgstate\_history json output ([issue#38846](#), [pr#27665](#), Brad Hubbard)
- osd: Include dups in copy\_after() and copy\_up\_to() ([issue#39304](#), [pr#27914](#), David Zafman)
- osd: Increase log level of messages which unnecessarily fill up logs ([pr#27686](#), David Zafman)
- osd: make osd recover more smoothly by avoiding failure peer info to resent ([pr#30404](#), xe5xaex8bxe9xa1xba10180185)
- osd: make PastIntervals a member of pg\_notify\_t ([pr#29517](#), Sage Weil)
- osd: merge replica log on primary need according to replica logs crt ([pr#29590](#), Zengran Zhang)
- osd: misc cleanups ([pr#30022](#), Yan Jun)
- osd: misc inc-recovery compat fixes ([pr#29754](#), xie xingguo)
- osd: optimize send\_message to peers ([pr#30968](#), Jianpeng Ma)
- osd: OSDMapRef access by multiple threads is unsafe ([pr#26874](#), Kefu Chai, Zengran Zhang)

- osd: Output Base64 encoding of CRC header if binary data present ([pr#27961](#), David Zafman)
- osd: partial recovery strategy based on PGLog ([pr#21722](#), lishuhao, Ning Yao)
- osd: peering updates peer\_last\_complete\_ondisk via setter ([pr#33659](#), Radoslaw Zarzynski)
- osd: pg as a mutex ([pr#29477](#), Kefu Chai)
- osd: prime splits/merges for any potential fabricated split/merge participant ([issue#38483](#), [pr#30018](#), xie xingguo)
- osd: process\_copy\_chunk remove obc ref before pg unlock ([issue#38842](#), [pr#27084](#), Zengran Zhang)
- osd: propagate mlcod to replicas and fix problems with read from replica ([pr#32381](#), Samuel Just, Sage Weil)
- osd: release backoffs during merge ([pr#31657](#), Sage Weil)
- osd: remove orphan include after PGLSParentFilter ([pr#29709](#), Radoslaw Zarzynski)
- osd: remove unused function ([pr#30644](#), Jianpeng Ma)
- osd: remove unused functions ([pr#32515](#), Jianpeng Ma)
- osd: Remove unused osdmap flags full, nearfull from output ([pr#30530](#), David Zafman)
- osd: remove useless ceph\_assert ([pr#31915](#), Jianpeng Ma)
- osd: revamp {noup,nodown,noin,noout} related commands ([pr#27735](#), xie xingguo)
- osd: rollforward may need to mark pglog dirty ([issue#36739](#), [pr#27015](#), Zengran Zhang)
- osd: scrub error on big objects; make bluestore refuse to start on big objects ([pr#29579](#), David Zafman, Sage Weil)
- osd: send smart asok result to stdout, not stderr ([pr#31412](#), Sage Weil)
- osd: set affinity for \\*all\\* threads ([pr#30712](#), Sage Weil)
- osd: set collection pool opts on collection create, pg load ([pr#29093](#), Sage Weil)
- osd: share curmap in handle\_osd\_ping ([pr#28662](#), Sage Weil)
- osd: shutdown recovery\_request\_timer earlier ([pr#27206](#), Zengran Zhang)
- osd: some prelim changes ([pr#29052](#), Sage Weil)
- osd: support osd\_repair\_during\_recovery ([issue#40620](#), [pr#28839](#), Jeegn Chen)

- osd: support osd\_scrub\_extended\_sleep ([issue#40955](#), [pr#29342](#), Jeegn Chen)
- osd: take heartbeat\_lock when calling heartbeat() ([issue#39439](#), [pr#27729](#), Sage Weil)
- osd: tiny clean-ups around the backfill ([pr#33583](#), Radoslaw Zarzynski)
- osd: track monotonic clock deltas between osds who ping each other ([pr#29116](#), Sage Weil, Samuel Just)
- osd: transpose two wait lists in comment ([pr#27017](#), Kefu Chai)
- osd: trim pg logs based on a per-osd budget ([pr#32683](#), Sage Weil, Kefu Chai)
- osd: Turn off repair pg state when leaving recovery ([pr#30852](#), David Zafman)
- osd: unify sources of no{up,down,in,out} flags into singleton helpers ([pr#28403](#), xie xingguo)
- osd: update comment as sub\_op\_scrub\_map has been removed ([pr#28338](#), Jing Wenjun)
- osd: Use physical ratio for nearfull (doesnt include backfill resserve) ([pr#31954](#), David Zafman)
- osd: use steady clock in prepare\_to\_stop() ([pr#26457](#), Mohamad Gebai)
- osd: use unique\_ptr for managing life cycles ([pr#32007](#), Kefu Chai)
- osdc/Striper: specialize std::min<> ([pr#28732](#), Kefu Chai)
- osd\_types: add ec profile to plain text osd pool ls detail output ([issue#40009](#), [pr#28224](#), Jan Fajerski)
- pybind,rbd: Add RBD\_FEATURE\_MIGRATING to rbd.pyx ([issue#39609](#), [pr#28009](#), Ricardo Marques)
- pybind,rbd: pybind/rbd: add config\_set/get/remove api in rbd.pyx ([pr#29459](#), Zheng Yin)
- pybind,rbd: pybind/rbd: add pool config\_set/get/remove api in rbd.pyx ([pr#30865](#), Zheng Yin)
- pybind,rbd: pybind/rbd: parent\_info should return pool namespace ([pr#30793](#), Ricardo Marques)
- pybind,rbd: rbd/pybind: fix unsupported format character of %lx ([pr#30314](#), songweibin)
- pybind,tests: pybind/rados: do not slice zip() ([pr#31044](#), Kefu Chai)
- pybind,tests: test/pybind/test\_rados.py: test test\_operate\_aio\_write\_op() ([pr#31158](#), Zhang Jiao)

- pybind/mgr: Add test\_orchestrator to mypy ([pr#32500](#), Sebastian Wagner)
- pybind/mgr: add\_tox\_test: Add mypy to TOX\_ENVS ([pr#32236](#), Sebastian Wagner)
- pybind/mgr: bump six to 1.14 ([pr#33185](#), Kefu Chai)
- pybind/tox: pass additional command line arguments through to tox ([pr#27947](#), Nathan Cutler)
- pybind: .gitignore: Add .mypy\_cache to .gitignore ([pr#33510](#), Kristoffer Grxc3xb6nlund)
- pybind: add verbose error message ([pr#28054](#), Daniel Badea, Changcheng Liu, Ovidiu Poncea)
- pybind: add WriteOp::set\_xattr() & rm\_xattr() ([pr#31829](#), Zhang Jiao)
- pybind: add writesame API ([pr#31489](#), Zhang Jiao)
- pybind: check CEPH\_LIBDIR not MAKEFLAGS ([pr#29080](#), Kefu Chai)
- pybind: customize compiler before checking cflags ([pr#33177](#), Kefu Chai)
- pybind: fix use of WriteOpCtx and ReadOpCtx ([issue#38946](#), [pr#27213](#), Ramana Raja)
- pybind: pybind/rados/rados.pyx: improve Rados.create\_pool() ([pr#31241](#), Zhang Jiao)
- pybind: pybind/rados: add application\_metadata\_get ([pr#30504](#), songweibin)
- pybind: pybind/rados: add Ioctx.get\_pool\_id() and Ioctx.get\_pool\_name() ([pr#29646](#), Zheng Yin)
- pybind: pybind/rados: add WriteOp::execute() ([pr#31546](#), Zhang Jiao)

- pybind: pybind/rados: should pass name to cstr() ([pr#27111](#), Kefu Chai)
- pybind: refactor monkey\_with\_compiler() ([pr#33061](#), Kefu Chai)
- pybind: set language\_level for cythonize explicitly ([pr#26607](#), Kefu Chai)
- python-common, mgr/orchestrator, mgr/dashboard: Use common Devices ([pr#30662](#), Kiefer Chang, Sebastian Wagner)
- python-common: add unmanaged property to PlacementSpec ([pr#33955](#), Sage Weil)
- python-common: all:true -> \\* ([pr#33970](#), Sage Weil)
- python-common: move pytest integration from setup.py to tox.ini ([pr#31943](#), Sebastian Wagner)
- python-common: remove all\_hosts from PlacementSpec ([pr#33948](#), Sebastian Wagner)
- qa/distros: rhel and centos: whitelist cephadm logrotate selinux denial ([pr#33110](#), Sage Weil)
- qa/standalone/test\_ceph\_daemon.sh: disable adoption for the moment ([pr#32178](#), Sage Weil)
- qa/standalone/test\_ceph\_daemon.sh: fix overwrites of temp files ([pr#31748](#), Sage Weil)
- qa/standalone/test\_ceph\_daemon: fix multi-version python test ([pr#31342](#), Sage Weil)
- qa/suites/cephadm: move orchestrator\_cli test into rados/cephadm ([pr#33648](#), Sage Weil)
- qa/suites/rados/ceph: drop opensuse for now ([pr#33801](#), Sage Weil)
- qa/suites/rados/cephadm/smoke: disable rgw role for now ([pr#33360](#), Sage Weil)
- qa/suites/rados/cephadm/upgrade: change start version ([pr#33475](#), Sage Weil)
- qa/suites/rados/cephadm/upgrade: fix initial version ([pr#33396](#), Sage Weil)
- qa/suites/rados/cephadm: explicitly test many distros ([pr#32969](#), Sage Weil)
- qa/suites/rados/cephadm: fix conflicts, missing .qa link ([pr#33132](#), Sage Weil)
- qa/suites/rados/cephadm[-smoke]: test podman on ubuntu 18.04 ([pr#33111](#), Sage Weil)
- qa/tasks/cephadm: ceph.git branches are now pushed to quay.io ([pr#32375](#), Sage Weil)
- qa/tasks/cephadm: deploy rgw daemons too ([pr#33289](#), Sage Weil)

- qa/tasks/cephadm: learn to pull cephadm from githu ([pr#32787](#), Sage Weil)
- qa/tasks/cephadm: misc fixes ([pr#32713](#), Sage Weil)
- qa/tasks/ceph\_manager.py: always use self.logger ([pr#29239](#), Kefu Chai)
- qa/tasks/ceph\_manager: 5s -> 15s for osd out to be visible ([pr#29013](#), Sage Weil)
- qa/tasks/ceph\_manager: fix movement of cot exports with cephadm ([pr#32986](#), Sage Weil)
- qa/tasks/ceph\_manager: fix shell osd for ceph-objectstore-tool commands ([pr#32725](#), Sage Weil)
- qa/tasks/ceph\_manager: make fix\_pgp\_num behave when no pool is found ([pr#32987](#), Sage Weil)
- qa/tasks/mgr/dashboard/test\_health: update schema ([pr#30507](#), Kefu Chai)
- qa/tasks/mgr/dashboard/test\_orchestrator: support addr attribute in inventory ([pr#33211](#), Kiefer Chang)
- qa/tasks/mgr/test\_orchestrator\_cli: fix device ls test ([pr#32384](#), Sage Weil)
- qa/tasks/mgr/test\_orchestrator\_cli: fix rgw add test ([pr#32101](#), Sage Weil)
- qa/tasks/mgr/test\_orchestrator\_cli: support multiple DriveGroups ([pr#33055](#), Kiefer Chang)
- qa/test: reduce over all number of runs ([pr#27979](#), Yuri Weinstein)
- qa/tests - cleaned up distro settings ([pr#27956](#), Yuri Weinstein)
- qa/tests - upped priority for upgrades on master, otherwise they nevexe2x80xa6 ([pr#29666](#), Yuri Weinstein)
- qa/tests: added nautilus-x-singleton suite to rados as symlink ([pr#27291](#), Sage Weil)
- qa/tests: added rados on master, reduced fs, rbd, multimds ([pr#27535](#), Yuri Weinstein)
- qa/tests: added the subset clause for nautilus branch ([pr#27129](#), Yuri Weinstein)
- qa/tests: changed the TO email to [ceph-qa@ceph.io](mailto:ceph-qa@ceph.io) ([pr#28721](#), Yuri Weinstein)
- qa/tests: moved some runs from ovh, removed ceph-disk/nautilus ([pr#27616](#), Yuri Weinstein)
- qa/tests: reduced runs for nautilus, added runs for octopus ([pr#33214](#), Yuri Weinstein)

- qa/tests: removed all runs on ovh ([pr#27960](#), Yuri Weinstein)
- qa/tests: removed filters for client-upgrade-\\* suites ([pr#28271](#), Yuri Weinstein)
- qa/tests: run luminous-x and mimic-x 2 times a week but with high priority ([pr#27527](#), Yuri Weinstein)
- qa/tests: trying to fix syntax error that prevented mimic-x to be addxe2x80xa6 ([pr#31799](#), Yuri Weinstein)
- qa/valgrind.supp: abstract from ceph::buffers symbol versioning ([pr#33757](#), Radoslaw Zarzynski)
- qa/workunits/cephadm/test\_adoption: run as root ([pr#33485](#), Sage Weil)
- qa/workunits/cephadm/test\_cephadm.sh: consolidate wait loop logic ([pr#33544](#), Michael Fritch)
- qa/workunits/cephadm/test\_cephadm.sh: dump logs on exit ([pr#33634](#), Michael Fritch)
- qa/workunits/cephadm/test\_cephadm.sh: need -fsid always ([pr#32220](#), Sage Weil)
- qa/workunits/cephadm/test\_cephadm.sh: re-enable adopt tests ([pr#32244](#), Michael Fritch)
- qa/workunits/cephadm/test\_cephadm.sh: skip docker when service is disabled ([pr#33018](#), Michael Fritch)
- qa/workunits/cephadm/test\_cephadm.sh: use available pythons; test on ubuntu and centos ([pr#32333](#), Sage Weil)
- qa/workunits/cephadm/test\_cephadm: -skip-monitoring-stack ([pr#34013](#), Sage Weil)
- qa/workunits/cephadm/test\_cephadm: fix typo ([pr#33181](#), Sage Weil)
- qa/workunits/cephadm/test\_cephadm: workunit test cleanup ([pr#32625](#), Michael Fritch)
- qa/workunits/cephadm/test\_repos: dont try to use the refspec ([pr#33134](#), Sage Weil)
- qa/workunits/cephadm: separate out test\_adoption.sh; fix ([pr#33457](#), Sage Weil)
- qa: fixes ([pr#29361](#), Kefu Chai)
- qa: misc fixes for rados and py3 ([pr#32362](#), Sage Weil)
- qa: pin rgw/verify to 8.0 ([pr#32761](#), Ali Maredia)
- qa: Run flake8 on python2 and python3 ([pr#32222](#), Thomas Bechtold)

- qa: vstart\_runner fails because of string index out of range ([pr#28990](#), Volker Theile)
- rbd,tests: cls/rbd: add snapshot limit UINT64\_MAX test case ([pr#31350](#), Chen Pan)
- rbd,tests: cls/rbd: add snapshot\_add raise -ESTALE test case ([pr#31149](#), wonderpow)
- rbd,tests: journal: always shutdown JournalRecorder before destructing it ([pr#29501](#), Kefu Chai)
- rbd,tests: journal: fix flush by age and in-flight byte tracking ([pr#31392](#), Jason Dillaman)
- rbd,tests: mgr/dashboard: s/fsid/mirror\_uuid/ ([pr#33348](#), Kefu Chai)
- rbd,tests: qa/rbd: add cram-based snap diff test ([issue#39447](#), [pr#28346](#), Shyukri Shyukriev, Nathan Cutler)
- rbd,tests: qa/suites/krbd: run unmap subsuite with msgr1 only ([pr#31265](#), Ilya Dryomov)
- rbd,tests: qa/suites/rbd: add random distro selection to librbd tests ([pr#27577](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: added writearound cache test permutations ([issue#39386](#), [pr#27694](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: fix errant tab in yaml which is causing parsing failures ([pr#30942](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: fixed download path for Ubuntu Bionic ([pr#32408](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: removed OpenStack tempest test cases ([pr#33900](#), Jason Dillaman)
- rbd,tests: qa/tests: added rbd task on ec ([pr#29541](#), Yuri Weinstein)
- rbd,tests: qa/workunit/rbd: fixed QoS throughput unit parsing ([pr#32280](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: fix compare\_images and compare\_image\_snapshots ([pr#28524](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: fixed python interpreter for EL8 ([pr#32409](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: fixups for the new krbd discard behavior ([pr#27192](#), Ilya Dryomov)
- rbd,tests: qa/workunits/rbd: override CEPH\_ARGS when initializing the site name

([pr#33187](#), Jason Dillaman)

- rbd,tests: qa/workunits/rbd: remove fast-diff from dynamic features test ([issue#39946](#), [pr#28135](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: stress test rbd mirror pool status -verbose ([pr#29655](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: use context managers to control Rados lifespan ([pr#34035](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: use https protocol for devstack git operations ([issue#39656](#), [pr#28063](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: use more recent qemu-io tests that support Bionic ([issue#24668](#), [pr#27683](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: wait for nbd map to close after unmap ([pr#33898](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: wait for rbd-nbd unmap to complete ([issue#39598](#), [pr#27981](#), Jason Dillaman)
- rbd,tests: qa: add device mapper and lvm test cases for stable pages ([pr#27271](#), Ilya Dryomov)
- rbd,tests: qa: add krbd\_discard\_granularity.t test ([pr#27042](#), Ilya Dryomov)
- rbd,tests: qa: add RBD QOS functional test ([pr#27137](#), Mykola Golub)
- rbd,tests: qa: add script to test how libceph handles huge osdmaps ([pr#30363](#), Ilya Dryomov)
- rbd,tests: qa: avoid hexdump skip and length options ([pr#30502](#), Ilya Dryomov)
- rbd,tests: qa: avoid page cache for krbd discard round off tests ([pr#30452](#), Ilya Dryomov)
- rbd,tests: qa: krbd\_parent\_overlap.t: fix read test ([pr#29966](#), Ilya Dryomov)
- rbd,tests: test/cli-integration/rbd: fixed missing image and snap ids ([pr#29853](#), Jason Dillaman)
- rbd,tests: test/cli-integration: fixed spacing issue for RBD formatted tables ([pr#33902](#), Jason Dillaman)
- rbd,tests: test/cls\_rbd/test\_cls\_rbd: update TestClsRbd.sparsify ([pr#30258](#), Kefu Chai)
- rbd,tests: test/cls\_rbd: include compat.h for ERESTART ([pr#32172](#), Willem Jan Withagen)

- rbd,tests: test/journal: always close object ([pr#29476](#), Kefu Chai)
- rbd,tests: test/librados\_test\_stub: ensure the log flusher thread is started ([pr#27326](#), Jason Dillaman)
- rbd,tests: test/librbd: allow parallel runs of run-rbd-unit-tests ([pr#30072](#), Willem Jan Withagen)
- rbd,tests: test/librbd: drop ceph\_test\_librbd\_api target ([issue#39072](#), [pr#27695](#), Jason Dillaman)
- rbd,tests: test/librbd: fix mock warnings in TestMockIoImageRequest ([pr#31497](#), Mykola Golub)
- rbd,tests: test/librbd: set nbd timeout due to newer kernels defaulting it on ([pr#29858](#), Jason Dillaman)
- rbd,tests: test/pybind/rbd.pyx: add test\_remove\_snap\_by\_id case in test\_rbd.py ([pr#30927](#), Zhang Jiao)
- rbd,tests: test/pybind: add create\_snap rasie ImageExists test case ([pr#31140](#), Gangbiao Liu)
- rbd,tests: test/pybind: inconsistent use of tabs and spaces in indentation ([pr#31606](#), Mykola Golub)
- rbd,tests: test/rbd\_mirror: fix mock warnings ([pr#31608](#), Mykola Golub)
- rbd,tests: test/run-rbd-tests: properly initialize newly created rbd pool ([pr#33642](#), Mykola Golub)
- rbd,tests: test: add test\_remove\_snap\_ImageNotFound test case in remove snap part ([pr#31221](#), Yingze Wei)
- rbd,tests: test:add test\_remove\_snap2 interface to remove snap when its protected ([pr#31208](#), Yingze Wei)
- rbd,tools: tools/rbd-ggate: close log before running postfork ([pr#30010](#), Willem Jan Withagen)
- rbd,tools: tools/rbd\_nbd: use POSIX basename() ([pr#28856](#), Kefu Chai)
- rbd-ggate: fix fallout from bufferlist.copy() change ([pr#33057](#), Willem Jan Withagen)
- rbd-mirror: add namespace support ([issue#37529](#), [pr#28939](#), Mykola Golub)
- rbd-mirror: add namespace support to service daemon ([pr#31642](#), Mykola Golub)
- rbd-mirror: add support for snapshot-based mirroring resyncs ([pr#33490](#), Jason Dillaman)

- rbd-mirror: apply image state during snapshot replay ([pr#33335](#), Jason Dillaman)
- rbd-mirror: cannot restore deferred deletion mirrored images ([pr#30351](#), Jason Dillaman)
- rbd-mirror: clear out bufferlist prior to listing mirror images ([issue#39407](#), [pr#27720](#), Jason Dillaman)
- rbd-mirror: continue to isolate journal replay logic ([pr#32399](#), Jason Dillaman)
- rbd-mirror: do not auto-create peers in non-default namespaces ([pr#32341](#), Jason Dillaman)
- rbd-mirror: dont expect image map is always initialized ([pr#33368](#), Mykola Golub)
- rbd-mirror: dont overwrite status error returned by replay ([pr#28179](#), Mykola Golub)
- rbd-mirror: ensure deterministic ordering of method calls ([pr#32274](#), Jason Dillaman)
- rbd-mirror: extract journal replaying logic from image replayer ([pr#32257](#), Jason Dillaman)
- rbd-mirror: fix pool replayer status for case when init failed ([pr#32483](#), Mykola Golub)
- rbd-mirror: fix race on namespace replayer initialization failure ([pr#32243](#), Mykola Golub)
- rbd-mirror: handle duplicates in image sync throttler queue ([issue#40519](#), [pr#28730](#), Mykola Golub)
- rbd-mirror: hold lock while updating local image name ([pr#33988](#), Jason Dillaman)
- rbd-mirror: ignore errors relating to parsing the cluster config file ([pr#29808](#), Jason Dillaman)
- rbd-mirror: image status should report remote status ([pr#30558](#), Jason Dillaman)
- rbd-mirror: improve detection of blacklisted state ([pr#33411](#), Mykola Golub)
- rbd-mirror: initial end-to-end test and associated bug fixes ([pr#33588](#), Jason Dillaman)
- rbd-mirror: initial snapshot replay state machine ([pr#33166](#), Jason Dillaman)
- rbd-mirror: initial snapshot-based mirroring bootstrap logic ([pr#33002](#), Jason Dillaman)
- rbd-mirror: link against the specified alloc library ([issue#40110](#), [pr#28434](#), Jason Dillaman)

- rbd-mirror: make logrotate work ([pr#32456](#), Mykola Golub)
- rbd-mirror: mirrored clone should be same format ([pr#31161](#), Mykola Golub)
- rbd-mirror: peer\_ping should send the local fsid to the remote ([pr#31950](#), Jason Dillaman)
- rbd-mirror: periodically flush IO and commit positions ([issue#39257](#), [pr#27533](#), Jason Dillaman)
- rbd-mirror: periodically poll remote mirror configuration ([pr#32671](#), Jason Dillaman)
- rbd-mirror: potential nullptr dereference in ImageReplayer::handle\_start\_replay ([pr#30484](#), Mykola Golub)
- rbd-mirror: prevent I/O modifications against a non-primary image ([pr#33831](#), Jason Dillaman)
- rbd-mirror: provide initial snapshot replay status ([pr#33440](#), Jason Dillaman)
- rbd-mirror: remove journal-specific logic from image replay and bootstrap state machines ([pr#32578](#), Jason Dillaman)
- rbd-mirror: removing non-primary trash snapshot ([pr#31260](#), Mykola Golub)
- rbd-mirror: rename per-image replication perf counters ([pr#32184](#), Mykola Golub)
- rbd-mirror: simplify peer bootstrapping ([pr#30411](#), Jason Dillaman)
- rbd-mirror: snapshot mirror mode ([pr#30548](#), Mykola Golub)
- rbd-mirror: snapshot-based mirroring should use image sync throttler ([pr#34040](#), Jason Dillaman)
- rbd-nbd: add netlink map/unmap support ([pr#27902](#), Mike Christie)
- rbd-nbd: add nl resize ([pr#29036](#), Mike Christie)
- rbd-nbd: sscnaf return 0 mean not-match ([issue#39269](#), [pr#27484](#), Jianpeng Ma)
- rbd: creating thick-provision image progress percent info exceeds 100% ([pr#30954](#), Xiangdong Mu)
- rbd: journal: add support for aligned appends ([pr#28351](#), Mykola Golub)
- rbd: librbd: skip stale child with non-existent pool for list descendants ([pr#29654](#), songweibin)
- rbd: add -merge to disk-usage ([pr#30994](#), Alexandre Bruyelles)
- rbd: add mirror snapshot schedule commands ([pr#32882](#), Mykola Golub)

- rbd: add snap\_exists method API ([pr#32497](#), Zheng Yin)
- rbd: client,common,mgr,rbd: clang related cleanups ([pr#33657](#), Kefu Chai)
- rbd: cls/rbd: improve efficiency of mirror image status queries ([pr#31865](#), Jason Dillaman)
- rbd: cls/rbd: sanitize entity instance messenger version type ([pr#30438](#), Jason Dillaman)
- rbd: cls/rbd: sanitize the mirror image status peer address after reading from disk ([pr#31824](#), Jason Dillaman)
- rbd: cls: reduce log level for non-fatal errors ([issue#40865](#), [pr#29165](#), Jason Dillaman)
- rbd: delete redundant words when trash restore fails because of same name ([pr#30952](#), Xiangdong Mu)
- rbd: fixed additional issues with CEPH\_ARGS processing ([pr#33219](#), Jason Dillaman)
- rbd: incorporate rbd-mirror daemon status in mirror pool status ([pr#31949](#), Jason Dillaman)
- rbd: journal: fix race between player shut down and cache rebalance ([pr#28748](#), Mykola Golub)
- rbd: journal: fix race between player shut down and cache rebalance ([pr#29796](#), Mykola Golub)
- rbd: journal: optimize object overflow detection ([pr#28240](#), Mykola Golub)
- rbd: journal: properly advance read offset after skipping invalid range ([pr#28627](#), Mykola Golub)
- rbd: journal: return error after first corruption detected ([pr#28820](#), Mykola Golub)
- rbd: journal: wait for in flight advance sets on stopping recorder ([pr#28529](#), Mykola Golub)
- rbd: krbd: avoid udev netlink socket overrun ([pr#30965](#), Ilya Dryomov)
- rbd: krbd: fix rbd map hang due to udev return subsystem unordered ([issue#39089](#), [pr#27339](#), Zhi Zhang)
- rbd: krbd: modprobe before calling build\_map\_buf() ([pr#30978](#), Ilya Dryomov)
- rbd: krbd: retry on transient errors from udev\_enumerate\_scan\_devices() ([pr#31023](#), Ilya Dryomov)
- rbd: krbd: return -ETIMEDOUT in polling ([issue#38792](#), [pr#27025](#), Dongsheng Yang)

- rbd: mgr/dashboard: support RBD mirroring bootstrap create/import ([issue#42355](#), [pr#31062](#), Jason Dillaman)
- rbd: msg/async: avoid unnecessary costly wakeups for outbound messages ([pr#28388](#), Jason Dillaman)
- rbd: msg/async: reduce verbosity of connection timeout failures ([issue#39448](#), [pr#28050](#), Jason Dillaman)
- rbd: pybind/mgr/rbd\_support: fix missing variable in error path ([pr#29773](#), Jason Dillaman)
- rbd: pybind/mgr/rbd\_support: ignore missing support for RBD namespaces ([pr#29433](#), Jason Dillaman)
- rbd: pybind/mgr/rbd\_support: use image ids to detect duplicate tasks ([pr#29468](#), Jason Dillaman)
- rbd: pybind/mgr/rbd\_support: wait for latest OSD map prior to handling commands ([pr#33451](#), Jason Dillaman)
- rbd: pybind/rbd: fix call to unregister\_osd\_perf\_queries ([pr#29419](#), Venky Shankar)
- rbd: pybind/rbd: provide snap remove flags ([pr#31627](#), Mykola Golub)
- rbd: qa/suites/rbd/openstack: use 18.04, not 16.04 ([pr#32284](#), Sage Weil)
- rbd: rbd-ggate: fix compile errors from ceph::mutex update ([pr#29474](#), Willem Jan Withagen)
- rbd: rbd-mirror: adjust journal fetch properties based on memory target ([pr#27670](#), Mykola Golub)
- rbd: rbd/action: display image id in rbd du/list output ([pr#29376](#), songweibin)
- rbd: rbd/action: fix error getting positional argument ([issue#40095](#), [pr#28313](#), songweibin)
- rbd: rbd/bench: outputs bytes/s format dynamically ([pr#31491](#), Zheng Yin)
- rbd: rbd/cache: Replicated Write Log core codes part 1 ([pr#31279](#), Peterson, Scott, Li, Xiaoyan, Lu, Yuan, Chamathy, Mahati)
- rbd: rbd/cache: Replicated Write Log core codes part 2 ([pr#31963](#), Peterson, Scott, Li, Xiaoyan, Lu, Yuan, Chamathy, Mahati)
- rbd: rbd\_replay: call the member decode() explicitly ([pr#27703](#), Kefu Chai)
- rbd: schedule for running trash purge operations ([pr#33389](#), Mykola Golub)
- rbd: src: use un-deprecated version of aio\_create\_completion ([pr#31333](#), Adam C.

Emerson)

- rbd: use the ordered throttle for the export action ([issue#40435](#), [pr#28657](#), Jason Dillaman)
- remove cephadm-adoption-corpus as submodule ([pr#33587](#), Sage Weil)
- Return an error, for Bluestore OSD, if WAL or DB are defined in the tags of the OSD but not present on the system ([pr#28791](#), David Casier)
- rgw,tests: qa/rgw/pubsub: fix tests to sync from master ([pr#33049](#), Yuval Lifshitz)
- rgw,tests: qa/rgw/pubsub: verify incremental sync is used in pubsu ([pr#33068](#), Yuval Lifshitz)
- rgw,tests: qa/rgw: add integration test for sse-kms with barbican ([pr#30218](#), Casey Bodley, Adam Kupczyk)
- rgw,tests: qa/rgw: add new rgw/website suite for static website tests ([pr#30193](#), Casey Bodley)
- rgw,tests: qa/rgw: add rgw\_obj and throttle tests to rgw verify suite ([pr#32188](#), Casey Bodley)
- rgw,tests: qa/rgw: disable debuginfo packages ([pr#27528](#), Casey Bodley)
- rgw,tests: qa/rgw: dont use ceph-ansible in s3a-hadoop suite ([issue#39706](#), [pr#28068](#), Casey Bodley)
- rgw,tests: qa/rgw: drop some objectstore types ([pr#30997](#), Casey Bodley)
- rgw,tests: qa/rgw: exercise DeleteRange in test\_bucket\_index\_log\_trim ([pr#33047](#), Casey Bodley)
- rgw,tests: qa/rgw: extra s3tests tasks use rgw endpoint configuration ([issue#17882](#), [pr#28631](#), Casey Bodley)
- rgw,tests: qa/rgw: fix import error in tasks/swift.py ([issue#40304](#), [pr#28605](#), Casey Bodley)
- rgw,tests: qa/rgw: fix swift warning message ([pr#28697](#), Casey Bodley)
- rgw,tests: qa/rgw: more fixes for swift task ([issue#40304](#), [pr#28823](#), Casey Bodley)
- rgw,tests: qa/rgw: multisite checkpoints consider pubsub zone ([pr#32941](#), Casey Bodley)
- rgw,tests: qa/rgw: refactor the kms backend configuration ([pr#30940](#), Casey Bodley)

- rgw, tests: qa/rgw: remove failing radosgw\_admin\_rest from multisite suite ([pr#32550](#), Casey Bodley)
- rgw, tests: qa/rgw: remove whitelist for SLOW\_OPS against ec pools ([pr#31363](#), Casey Bodley)
- rgw, tests: qa/rgw: s3a-hadoop task defaults to maven-version 3.6.3 ([pr#32620](#), Casey Bodley)
- rgw, tests: qa/rgw: skip swift tests on rhel 7.6+ ([issue#40304](#), [pr#28532](#), Casey Bodley)
- rgw, tests: qa/rgw: update run-s3tests.sh ([pr#28964](#), Casey Bodley)
- rgw, tests: qa/rgw: use testing kms backend for multisite tests ([pr#31374](#), Casey Bodley)
- rgw, tests: qa/rgw: use testing kms backend for other rgw subsuites ([pr#31414](#), Casey Bodley)
- rgw, tests: qa/rgw: whitelist SLOW\_OPS failures against ec pools ([pr#30944](#), Casey Bodley)
- rgw, tests: qa/suites/rgw/website: run test on ubuntu ([pr#32791](#), Sage Weil)
- rgw, tests: qa/suites/rgw: reenable ragweed (now py3) ([pr#32310](#), Sage Weil)
- rgw, tests: qa/suites: use s3-tests with python3 support ([pr#32624](#), Ali Maredia)
- rgw, tests: qa/tasks/swift: remove swift tests ([pr#32357](#), Sage Weil)
- rgw, tests: qa/tests: added rgw into upgrade sequence to improve coverage ([pr#29234](#), Yuri Weinstein)
- rgw, tests: qa/tests: added rgw into upgrade sequence to improve coverage - splits ([pr#29282](#), Yuri Weinstein)
- rgw, tests: qa: add force-branch to suites running s3readwrite & s3roundtrip tasks ([pr#32225](#), Ali Maredia)
- rgw, tests: qa: bump maven repo version in s3a\_hadoop.py ([pr#30531](#), Ali Maredia)
- rgw, tests: qa: radosgw-admin: remove dependency on bunch package ([pr#32100](#), Yehuda Sadeh)
- rgw, tests: qa: radosgw\_admin: validate a simple user stats output ([pr#30684](#), Abhishek Lekshmanan)
- rgw, tests: qa: remove mon valgrind check in rgw verify suite ([issue#38827](#), [pr#28155](#), Ali Maredia)
- rgw, tests: qa: remove s3-tests from rados/basic/tasks/rgw\_snaps.yml ([pr#32940](#),

Ali Maredia)

- rgw,tests: qa: rgw: add user-policy caps for the s3tests users ([pr#31127](#), Abhishek Lekshmanan)
- rgw,tests: qa: use curl in wait\_for\_radosgw() in util/rgw.py ([pr#28521](#), Ali Maredia)
- rgw,tests: rgw/amqp: fix race condition in AMQP unit test ([pr#30735](#), Yuval Lifshitz)
- rgw,tests: rgw/amqp: remove flaky amqp test ([pr#31510](#), Yuval Lifshitz)
- rgw,tests: rgw/pubsub: add multisite pubsub tests to teuthology ([pr#27838](#), Yuval Lifshitz)
- rgw,tests: rgw/pubsub: tests enhancements and fixes ([pr#28910](#), Yuval Lifshitz)
- rgw,tests: rgw/pubsub: use incremental sync for pubsub module by default ([pr#28470](#), Yuval Lifshitz)
- rgw,tests: test/rgw: fix test-rgw-multisite.sh script for creating multisite clusters ([pr#27984](#), Casey Bodley)
- rgw,tests: test/rgw: fixes for test-rgw-multisite.sh ([pr#33537](#), Casey Bodley)
- rgw,tests: test/rgw: raise timer durations for unittest\_rgw\_reshard\_wait ([pr#32094](#), Casey Bodley)
- rgw,tests: test/rgw: test\_rgw\_reshard\_wait uses same clock for timing ([pr#27035](#), Casey Bodley)
- rgw,tests: vstart: move common rgw config to [client.rgw] ([pr#29449](#), Casey Bodley)
- rgw,tools: ceph-dencoder: add RGWPeriodLatestEpochInfo support ([pr#30613](#), yuliyang)
- rgw,tools: rgw/examples: adding examples for boto3 extensions to AWS S3 ([pr#30600](#), Yuval Lifshitz)
- rgw,tools: vstart.sh: run multiple rgws with different ids ([pr#26690](#), Joao Eduardo Luis)
- rgw: rgw: cls\_bucket\_list\_unordered lists a single shard ([issue#39393](#), [pr#27697](#), Casey Bodley)
- rgw: rgw: make radosgw-admin user create and modify distinct ([pr#31901](#), Matthew Oliver)
- rgw: rgw: returns LimitExceeded when user creates too many ACLs ([issue#26835](#), [pr#25692](#), Chang Liu)

- rgw: A task to run S3 Java tests against RGW ([pr#22788](#), Antoaneta Damyanova)
- rgw: add -object-version in radosgw-admin help info ([pr#30091](#), yuliyang)
- rgw: add a small efficiency ([pr#29178](#), J. Eric Ivancich)
- rgw: add admin rest api for bucket sync ([pr#19020](#), zhang Shaowen, Zhang Shaowen)
- rgw: add cls\_queue and cls\_rgw\_gc for omap offload ([pr#28421](#), Pritha Srivastava, Casey Bodley)
- rgw: add const correctness to some rest functions ([pr#31660](#), J. Eric Ivancich)
- rgw: add creation time information into bucket stats ([pr#30384](#), Enming Zhang)
- rgw: Add days0 to rgw lc ([pr#29937](#), Or Friedmann)
- rgw: add detailed error message for PutACLs ([pr#30385](#), Enming Zhang)
- rgw: add editor directive comments to rgw services source files ([pr#27897](#), J. Eric Ivancich)
- rgw: add GET /admin/realm?list api to list realms ([pr#28156](#), Casey Bodley)
- rgw: add missing admin property when sync user info ([pr#30127](#), zhang Shaowen)
- rgw: add missing bilog status to help info ([pr#30357](#), zhang Shaowen)
- rgw: add missing close\_section in send\_versioned\_response ([pr#28946](#), Casey Bodley)
- rgw: Add more details to the LC delete and transit log ([pr#30913](#), Or Friedmann)
- rgw: add num\_shards to radosgw-admin bucket stats ([pr#30845](#), Paul Emmerich)
- rgw: add option to specify shard-id for bi list admin command ([pr#29394](#), Mark Kogan)
- rgw: add optional\_yield to http client interface ([pr#25355](#), Casey Bodley)
- rgw: add optional\_yield to SysObj service interfaces ([pr#25353](#), Casey Bodley)
- rgw: add PublicAccessBlock set of APIs on buckets ([pr#30033](#), Abhishek Lekshmanan)
- rgw: add rgw\_rados\_pool\_recovery\_priority (default 5) ([pr#29181](#), Sage Weil)
- rgw: add roles\_pool in RGWZoneParams dump/decode json ([issue#22162](#), [pr#17338](#), Tianshan Qu)
- rgw: add S3 object lock feature to support object worm ([pr#26538](#), zhang Shaowen)
- rgw: add some comments to rgw code to help explain functionality ([pr#27896](#), J. Eric Ivancich)

- rgw: add SSE-KMS with Vault using token auth ([pr#29783](#), Andrea Baglioni, Sergio de Carvalho)
- rgw: Add support bucket policy for subuser ([pr#33165](#), Seena Fallah)
- rgw: add tenant as parameter to User in multisite tests ([pr#27969](#), Yuval Lifshitz)
- rgw: add transaction id to ops log ([pr#30163](#), zhang Shaowen)
- rgw: add YieldingAioThrottle for async PutObj/GetObj ([pr#26173](#), Casey Bodley)
- rgw: Added caching for S3 credentials retrieved from keystone ([pr#26095](#), James Weaver)
- rgw: adding documentation for AssumeRoleWithWebIdentity ([pr#31994](#), Pritha Srivastava)
- rgw: Adding iam namespace for Role and User Policy related REST APIs ([pr#27178](#), Pritha Srivastava)
- rgw: adding mfa code validation when bucket versioning status is changed ([pr#31767](#), Pritha Srivastava)
- rgw: Adding tcp\_nodelay option to Beast ([pr#27008](#), Or Friedmann)
- rgw: address 0-length listing results when non-vis entries dominate ([pr#32636](#), J. Eric Ivancich)
- rgw: adjust allowable bucket index shard counts for dynamic resharding ([pr#30795](#), J. Eric Ivancich)
- rgw: admin: handle delete\_at attr in object stat output ([pr#27781](#), Abhishek Lekshmanan)
- rgw: Allow admin APIs that write metadata to be executed first on the mastxe2x80xa6 ([issue#39549](#), [pr#29549](#), Shilpa Jagannath)
- rgw: allow radosgw-admin to list bucket w -allow-unordered ([issue#39637](#), [pr#28031](#), J. Eric Ivancich)
- rgw: allow reshards log entries for non-existent buckets to be cancelled ([pr#31271](#), J. Eric Ivancich)
- rgw: apply\_olh\_log ignores RGW\_ATTR\_OLH\_VER decode error ([pr#31976](#), Casey Bodley)
- rgw: asio: check the remote endpoint before processing requests ([pr#29967](#), Abhishek Lekshmanan)
- rgw: auth/Crypto: fallback to /dev/urandom if getentropy() fails ([pr#30544](#), Kefu Chai)

- rgw: auto-clean reshards queue entries for non-existent buckets ([pr#31323](#), J. Eric Ivancich)
- rgw: az: add archive zone tests ([pr#29359](#), Javier M. Mellid)
- rgw: beast frontend uses 512k mprotected coroutine stacks ([pr#31580](#), Daniel Gryniewicz, Casey Bodley)
- rgw: beast frontend uses yield\_context to read/write body ([pr#27795](#), Casey Bodley)
- rgw: beast port parsing ([issue#39000](#), [pr#27242](#), Abhishek Lekshmanan)
- rgw: beast ssl certs config through config-key ([pr#33287](#), Yehuda Sadeh)
- rgw: bucket granularity sync ([pr#31686](#), Yehuda Sadeh)
- rgw: bucket re-creation fixes ([pr#32121](#), Yehuda Sadeh)
- rgw: bucket stats report mtime in UTC ([pr#27617](#), Casey Bodley)
- rgw: bucket tagging ([pr#27993](#), Chang Liu)
- rgw: build async scheduler only when beast is built ([pr#26634](#), Abhishek Lekshmanan)
- rgw: build radosgw daemon as a shared lib + small executable ([pr#32404](#), Kaleb S. Keithley)
- rgw: build\_linked\_oids\_for\_bucket and build\_buckets\_instance\_index should return negative value if it fails ([pr#31346](#), zhangshaowen)
- rgw: change cls rgw reshards status to enum class ([pr#30611](#), J. Eric Ivancich)
- rgw: change MAX\_USAGE\_TRIM\_ENTRIES value from 128 to 1000 ([pr#30392](#), zhang Shaowen)
- rgw: check lc objs not empty after fetching ([pr#26167](#), Yao Zongyou)
- rgw: clean index and remove bucket instance info when setting resharding status fails ([pr#31103](#), zhangshaowen)
- rgw: clean up ordered list ([pr#31338](#), J. Eric Ivancich)
- rgw: clean up some logging ([pr#27411](#), J. Eric Ivancich)
- rgw: cleanup the magic string usage in `cls_rgw_client.cc` ([pr#31432](#), zhangshaowen)
- rgw: cleanup:remove un-used class member in `RGWDeleteLC` ([pr#31404](#), zhang Shaowen)
- rgw: cleanup:remove un-used `create_new_bucket_instance` in `rgw_admin.cc` ([pr#31345](#), zhangshaowen)

- rgw: clear ent\_list for each loop of bucket list ([issue#44394](#), [pr#33693](#), Yao Zongyou)
- rgw: cls/rgw: fix bilog trim tests in ceph\_test\_cls\_rgw ([pr#30268](#), Casey Bodley)
- rgw: cls/rgw: keep issuing bilog trim ops after reset ([issue#40187](#), [pr#28430](#), Casey Bodley)
- rgw: cls/rgw: test before accessing pkeys->rbegin() ([issue#39984](#), [pr#28391](#), Casey Bodley)
- rgw: cls/rgw: when object is versioned and lc transition it, the object is becoming non-current ([pr#32458](#), Or Friedmann)
- rgw: cls/user: cls\_user\_set\_buckets\_info overwrites creation\_time ([issue#39635](#), [pr#28045](#), Casey Bodley)
- rgw: cls\_bucket\_list\\_(un)ordered should clear results collection ([pr#33702](#), J. Eric Ivancich)
- rgw: compression info should be same during multipart uploading ([pr#30574](#), zhang Shaowen)
- rgw: conditionally allow non-unique email addresses ([issue#40089](#), [pr#28327](#), Matt Benjamin)
- rgw: continuation token doesnt work in list object v2 request ([pr#28988](#), zhang Shaowen)
- rgw: continuationToken or startAfter shouldnt be returned if not specified ([pr#29298](#), zhang Shaowen)
- rgw: correct some error log about reshards in cls\_rgw.cc ([pr#31429](#), zhangshaowen)
- rgw: crypt: permit RGW-AUTO/default with SSE-S3 headers ([pr#30189](#), Matt Benjamin)
- rgw: crypto: throw DigestException from Digest and HMAC ([issue#39456](#), [pr#27765](#), Matt Benjamin)
- rgw: data sync markers include timestamp from datalog entry ([pr#32309](#), Casey Bodley)
- rgw: data/bilogs are trimmed when no peers are reading them ([issue#39487](#), [pr#27794](#), Casey Bodley)
- rgw: datalog/mdlog trim commands loop until done ([pr#29448](#), Casey Bodley)
- rgw: data\_sync\_source\_zones only contains exporting zones ([pr#33193](#), Casey Bodley)
- rgw: decrypt filter does not cross multipart boundaries ([issue#38700](#), [pr#27130](#), Adam Kupczyk, Casey Bodley, Abhishek Lekshmanan)

- rgw: DefaultRetention requires either Days or Years ([pr#29680](#), Chang Liu)
- rgw: delete\_obj\_index() takes mtime for bilog ([issue#24991](#), [pr#27980](#), Casey Bodley)
- rgw: distinguish different get\_usage for usage log ([pr#17719](#), Jiaying Ren)
- rgw: dmclock: wait until the request is handled ([pr#30777](#), GaryHyyg)
- rgw: do not miss the 1000th element of every iteration during lifecycle processing ([pr#30861](#), Ilsoo Byun)
- rgw: do not remove delete marker when fixing versioned bucket ([pr#32562](#), Ilsoo Byun)
- rgw: Dont crash on copy when metadata directive not supplied ([issue#40416](#), [pr#28949](#), Adam C. Emerson)
- rgw: dont crash on missing /etc/mime.types ([issue#38328](#), [pr#26998](#), Casey Bodley)
- rgw: dont print error log when list reshards result is not truncated ([pr#31142](#), zhangshaowen)
- rgw: dont recalculate etags for slo/dlo ([pr#27470](#), Casey Bodley)
- rgw: dont throw when accept errors are happening on frontend ([pr#29587](#), Yuval Lifshitz)
- rgw: drop cloud sync module logs attrs from the log ([pr#27820](#), Nathan Cutler)
- rgw: drop dead flush\_read\_list declaration ([pr#29458](#), Jiaying Ren)
- rgw: drop unused rgw\_decode\_pki\_token() ([pr#27052](#), Radoslaw Zarzynski)
- rgw: dump s3\_code as the Code response element in RGWDeleteMultiObj\_ObjStore\_S3 ([issue#18241](#), [pr#12470](#), Radoslaw Zarzynski)
- rgw: eliminates duplicated tags\_b1 var ([pr#27970](#), Chang Liu)
- rgw: Evaluating bucket policies also while reading permissions for anxe2x80xa6 ([issue#38638](#), [pr#27309](#), Pritha Srivastava)
- rgw: examples: rgw: add boto3 append & get usage api extensions ([pr#33063](#), Abhishek Lekshmanan)
- rgw: Expiration days cant be zero and transition days can be zero ([pr#30878](#), zhang Shaowen)
- rgw: extend SSE-KMS with Vault using transit secrets engine ([pr#31361](#), Andrea Baglioni, Sergio de Carvalho)
- rgw: fetch\_remote\_obj() compares expected object size ([pr#28303](#), Xiaoxi CHEN,

Casey Bodley)

- rgw: find oldest period and update RGWMetadataLogHistory() ([pr#31873](#), Shilpa Jagannath)
- rgw: fix a bug that bucket instance obj cant be removed after resharding completed ([pr#31483](#), zhang Shaowen)
- rgw: fix a bug that lifecycle expiraton generates delete marker continuously ([issue#40393](#), [pr#28587](#), zhang Shaowen)
- rgw: fix bucket may redundantly list keys after BI\_PREFIX\_CHAR ([issue#39984](#), [pr#28188](#), Tianshan Qu)
- rgw: Fix bucket versioning vs. swift metadata bug ([pr#29240](#), Marcus Watts)
- rgw: Fix bug on subuser policy identity checker ([pr#33398](#), Seena Fallah)
- rgw: fix bug with (un)ordered bucket listing and marker w/ namespace ([pr#33046](#), J. Eric Ivancich)
- rgw: fix bugs in listobjectsv1 ([pr#28873](#), Albin Antony)
- rgw: fix cls\_bucket\_list\_unordered() partial results ([pr#29692](#), Mark Kogan)
- rgw: fix compile errors with boost 1.70 ([pr#27730](#), Casey Bodley)
- rgw: fix data consistency error casued by rgw sent timeout ([pr#30257](#), xe6x9dx8exe7xbaxb2xe5xbdxac82225)
- rgw: fix data sync start delay if remote havent init data\_log ([pr#30393](#), Tianshan Qu)
- rgw: fix default storage class for get\_compression\_type ([pr#29909](#), Casey Bodley)
- rgw: fix default\_placement containing / when storage\_class is standard ([issue#39380](#), [pr#27676](#), mkogan1)
- rgw: fix dns name comparison for virtual hosting ([pr#30221](#), Casey Bodley)
- rgw: Fix documentation for rgw\_ldap\_secret ([pr#29816](#), Robin Mxc3xbcller)
- rgw: fix drain handles error when deleting bucket with bypass-gc option ([pr#28789](#), dongdong tao)
- rgw: Fix dynamic resharding not working for empty zonegroup in period ([pr#31977](#), Or Friedmann)
- rgw: Fix expiration header does not return the earliest rule ([pr#29399](#), Or Friedmann)
- rgw: fix incorrect radosgw-admin zonegroup rm info ([pr#30319](#), zhang Shaowen)

- rgw: fix indentation for listobjects v2 ([pr#28830](#), Albin Antony)
- rgw: fix list bucket with delimiter wrongly skip some special keys ([issue#40905](#), [pr#29215](#), Tianshan Qu)
- rgw: fix list bucket with start marker and delimiter / will miss next object `xe2x80xa6` ([issue#39989](#), [pr#28192](#), Tianshan Qu)
- rgw: fix list versions starts with version\_id=null ([pr#29897](#), Tianshan Qu)
- rgw: fix MalformedXML errors in PutBucketObjectLock/PutObjRetention ([pr#28783](#), Casey Bodley)
- rgw: fix memory growth while deleting objects with ([pr#30174](#), Mark Kogan)
- rgw: fix minimum of unordered bucket listing ([pr#30146](#), J. Eric Ivancich)
- rgw: fix minor compiler warning in keystone auth ([pr#27100](#), David Disseldorp)
- rgw: fix miss get ret in STSService::storeARN ([issue#40386](#), [pr#28527](#), Tianshan Qu)
- rgw: fix miss handle curl error return ([pr#28345](#), Casey Bodley, Tianshan Qu)
- rgw: fix missing tenant prefix in bucket name during bucket link ([pr#29815](#), Shilpa Jagannath)
- rgw: fix multipart uploads error response ([pr#32771](#), GaryHyg)
- rgw: Fix narrowing conversion error ([pr#28905](#), Adam C. Emerson)
- rgw: fix one part of the bulk delete(RGWDeleteMultiObj\_ObjStore\_S3) fails but no error messages ([pr#29795](#), Snow Si)
- rgw: fix opslog operation field as per Amazon s3 ([issue#20978](#), [pr#30539](#), Jiaying Ren)
- rgw: fix potential realm watch lost ([issue#40991](#), [pr#29369](#), Tianshan Qu)
- rgw: fix read not exists null version return wrong ([issue#38811](#), [pr#27047](#), Tianshan Qu)
- rgw: fix refcount tags to match and update objects idtag ([pr#30013](#), J. Eric Ivancich)
- rgw: fix REQUEST\_URI setting in the rgw\_asio\_client.cc ([pr#30540](#), Jiaying Ren)
- rgw: fix rgw crash and set correct error code ([pr#28172](#), yuliyang)
- rgw: fix rgw crash when duration is invalid in sts request ([pr#32119](#), yuliyang)
- rgw: fix rgw crash when token is not base64 encode ([pr#31830](#), yuliyang)

- rgw: fix rgw decompression log-print ([pr#29633](#), Han Fengzhe)
- rgw: fix rgw lc does not delete objects that do not have exactly the same tags as the rule ([pr#30151](#), Or Friedmann)
- rgw: fix RGWDeleteMultiObj::verify\_permission() ([pr#26947](#), Irek Fasikhov)
- rgw: fix RGWUserInfo decode current version ([pr#31591](#), Chang Liu)
- rgw: fix S3 compatibility bug when CORS is not found ([issue#37945](#), [pr#25999](#), Nick Janus)
- rgw: fix sharded bucket listing with prefix/delimiter ([pr#33628](#), Casey Bodley)
- rgw: fix SignatureDoesNotMatch when use ipv6 address in s3 client ([pr#30778](#), yuliyang)
- rgw: fix signed char truncation in delimiter check ([pr#27001](#), Matt Benjamin)
- rgw: fix string\_view formatting in RGWFormatter\_Plain ([pr#33754](#), Casey Bodley)
- rgw: fix the bug of rgw not doing necessary checking to website configuration ([issue#40678](#), [pr#28904](#), Enming Zhang)
- rgw: fix unlock of shared lock in RGWCache ([pr#29558](#), Abhishek Lekshmanan)
- rgw: fix unlock of shared lock in RGWDataChangesLog ([pr#29538](#), Casey Bodley)
- rgw: Fix upload part copy range able to get almost any string ([pr#32487](#), Or Friedmann)
- rgw: fix version tracking across bucket link steps ([pr#29851](#), Matt Benjamin)
- rgw: fixed unrecognized arg error when using radosgw-admin zone rm ([pr#30060](#), Hongang Chen)
- rgw: Fixes related to omap offload and gc ([pr#33372](#), Pritha Srivastava)
- rgw: followup for user rename ([pr#29540](#), Casey Bodley)
- rgw: forwarded some requests to master zone ([pr#28276](#), Chang Liu)
- rgw: gc remove tag after all sub io finish ([issue#40903](#), [pr#29199](#), Tianshan Qu)
- rgw: get barbican secret key request maybe return error code ([pr#29639](#), Richard Bai(xe7x99xbdx5xadxa6xe4xbdx99))
- rgw: get elastic search info in start\_sync, avoid creating new coroutines manager ([pr#32269](#), Chang Liu)
- rgw: housekeeping of reset stats operation in radosgw-admin and cls back-end ([pr#29515](#), J. Eric Ivancich)

- rgw: http client drops lock before suspending coroutine ([pr#29553](#), Casey Bodley)
- rgw: iam: add all http args to req\_info ([pr#31124](#), Abhishek Lekshmanan)
- rgw: iam: use a function to calculate the Action Bit string ([pr#30152](#), Abhishek Lekshmanan)
- rgw: ignore If-Unmodified-Since if If-Match exists, and ignore If-Modified-Since if If-None-Match exists ([pr#28625](#), zhang Shaowen)
- rgw: improve beast ([pr#33017](#), Or Friedmann, Matt Benjamin)
- rgw: improve data sync restart after failure ([pr#30175](#), Tianshan Qu)
- rgw: improve debugs on the path of RGWRados::cls\_bucket\_head ([pr#12709](#), Radoslaw Zarzynski)
- rgw: improvements to SSE-KMS with Vault ([pr#31025](#), Andrea Baglioni, Sergio de Carvalho)
- rgw: Improving doc for Cross Project(Tenant) access with Openstack Kexe2x80xa6 ([pr#27507](#), Pritha Srivastava)
- rgw: incorrect return value when processing CORS headers ([pr#28622](#), Ilsoo Byun)
- rgw: Incorrectly calling ceph::buffer::list::decode\_base64 in bucket policy ([pr#31356](#), GaryHyg)
- rgw: increase beast parse buffer size to 64k ([pr#29776](#), Casey Bodley)
- rgw: increase log level for same or older period pull msg ([pr#33527](#), Ali Maredia)
- rgw: Increase the default number of RGW bucket shards ([pr#32660](#), Casey Bodley, Mark Nelson)
- rgw: init-radosgw: use ceph-conf to get cluster configuration value ([pr#27538](#), Daniel Badea)
- rgw: Initialize member variables in rgw\_sync.h, rgw\_rados.h ([pr#16929](#), amitkuma)
- rgw: initialize member variables of rgw\_log\_entry ([pr#32430](#), Kefu Chai)
- rgw: kill compile warnning in rgw\_object\_lock.h ([pr#30489](#), Chang Liu)
- rgw: LC expiration header should present midnight expiration date ([pr#31887](#), Or Friedmann)
- rgw: lc: check for valid placement target before processing transitions ([pr#28256](#), Abhishek Lekshmanan)
- rgw: LC: handle resharded buckets ([pr#26564](#), Abhishek Lekshmanan)

- rgw: ldap auth: S3 auth failure should return InvalidAccessKeyId ([pr#30332](#), Matt Benjamin)
- rgw: ldap: fix LDAPAuthEngine::init() when uri !empty() ([pr#26911](#), Matt Benjamin)
- rgw: lifecycle days may be 0 ([pr#26524](#), Matt Benjamin)
- rgw: lifecycle: alternate solution to prefix\_map conflict ([issue#37879](#), [pr#26518](#), Matt Benjamin)
- rgw: limit entries in remove\_olh\_pending\_entries() ([issue#39118](#), [pr#27400](#), Casey Bodley)
- rgw: list buckets: dont return buckets if limit=0 ([pr#32109](#), Yehuda Sadeh)
- rgw: list\_bucket versions return NextVersionIdMarker = null if next\_marker.instance is empty ([pr#17591](#), Shasha Lu)
- rgw: log refactoring for putobj\_processor ([pr#26107](#), Ali Maredia)
- rgw: log refactoring for rgw\_rest\_s3/swift ops ([pr#27037](#), Ali Maredia)
- rgw: make dns hostnames matching case insensitive ([issue#40995](#), [pr#29380](#), Abhishek Lekshmanan)
- rgw: make max\_connections configurable in beast ([pr#33053](#), Tiago Pasqualini)
- rgw: Make rgw admin ops api get user info consistent with the command line ([pr#26183](#), Li Shuhao)
- rgw: make sure modelines are correct for all files ([pr#29742](#), Daniel Gryniewicz)
- rgw: maybe coredump when reload operator happened ([pr#29733](#), Richard Bai(xe7x99xbdx5xadxa6xe4xbdx99))
- rgw: metadata refactoring ([pr#29118](#), Casey Bodley, Yehuda Sadeh)
- rgw: mgr/ansible: Change default realm and zonegroup ([pr#29793](#), Sebastian Wagner)
- rgw: mgr/dashboard: enable/disable MFA Delete on RGW bucket ([pr#31922](#), Alfonso Martxc3xadnez)
- rgw: mgr/orchestrator: name rgw by client.rgw.\$realm.\$zone[\$id] ([pr#31890](#), Sage Weil)
- rgw: mitigate bucket list with max-entries excessively high ([pr#29179](#), J. Eric Ivancich)
- rgw: move bucket reshards checks out of write path ([pr#29852](#), Casey Bodley)
- rgw: move delimiter-based bucket listing/filtering logic to cls ([pr#30272](#), J. Eric Ivancich)

- rgw: move forward marker even in case of many rgw.none indexes ([pr#32513](#), Ilsoo Byun)
- rgw: Move upload\_info declaration out of conditional ([pr#29559](#), Adam C. Emerson)
- rgw: multipart upload abort is best-effort ([issue#40526](#), [pr#28724](#), J. Eric Ivancich)
- rgw: MultipartObjectProcessor supports stripe size > chunk size ([pr#32996](#), Casey Bodley)
- rgw: multisite log trimming only checks peers that sync from us ([issue#39283](#), [pr#27567](#), Casey Bodley)
- rgw: nfs: skip empty (non-POSIX) path segments ([issue#38744](#), [pr#26954](#), Matt Benjamin)
- rgw: nfs: svc-enable RGWLi ([pr#26981](#), Matt Benjamin)
- rgw: normalize v6 endpoint behaviour for the beast frontend ([issue#39038](#), [pr#27270](#), Abhishek Lekshmanan)
- rgw: object expirer fixes ([pr#27870](#), Abhishek Lekshmanan)
- rgw: Object tags shouldnt work with deletemarker or multipart expiration ([issue#40405](#), [pr#28617](#), zhang Shaowen)
- rgw: one log shard fails shouldnt block other shards process when reshards buckets ([pr#31155](#), zhangshaowen)
- rgw: One Rados Handle to Rule Them All ([pr#27102](#), Adam C. Emerson)
- rgw: orphan fixes ([pr#26412](#), Abhishek Lekshmanan)
- rgw: parse\_copy\_location defers url-decode ([issue#27217](#), [pr#25498](#), Casey Bodley)
- rgw: perfcounters: add gc retire counter ([pr#26351](#), Matt Benjamin)
- rgw: permit rgw-admin to populate user info by access-key ([pr#28331](#), Matt Benjamin)
- rgw: Policy should be url\_decode when assume\_role ([pr#28704](#), yuliyang)
- rgw: prefix-delimiter listing: support >1 character delimiter ([pr#26863](#), Matt Benjamin)
- rgw: prevent bucket reshards scheduling if bucket is resharding ([pr#30610](#), J. Eric Ivancich)
- rgw: prevent LC from reading stale head when transitioning object ([pr#31214](#), Ilsoo Byun)

- rgw: project and return lc expiration from GET/HEAD and PUT ops ([pr#26160](#), Matt Benjamin)
- rgw: Project Zipper - Bucket ([pr#31436](#), Daniel Gryniewicz)
- rgw: Project Zipper - Bucketlist ([pr#30619](#), Daniel Gryniewicz)
- rgw: Project Zipper part 1 ([pr#28824](#), Daniel Gryniewicz)
- rgw: qa/suite/rgw/verify: valgrind on centos again! ([pr#32727](#), Sage Weil)
- rgw: qa/tasks/s3tests\_java: move to gradle 6.0.1 ([pr#32335](#), Sage Weil)
- rgw: qa/tests: update s3a hadoop versions used for test ([pr#26100](#), Vasu Kulkarni)
- rgw: qa: remove force-branch from overrides of s3-tests ([pr#32462](#), Ali Maredia)
- rgw: qa: update s3-test download code for s3-test tasks ([pr#31839](#), Ali Maredia)
- rgw: queue like an Egyptian([pr#26461](#), Adam C. Emerson)
- rgw: race condition between resharding and ops waiting on resharding ([issue#38990](#), [pr#27223](#), J. Eric Ivancich)
- rgw: radosgw-admin flush user stats output ([pr#30669](#), Abhishek Lekshmanan)
- rgw: radosgw-admin zone placement rm and radosgw-admin zonegroup placement rm support -storage-class ([pr#31239](#), yuliyang)
- rgw: radosgw-admin: add -uid check in bucket list command ([pr#30194](#), Vikhyat Umrao)
- rgw: radosgw-admin: bucket sync status not caught up during full sync ([issue#40806](#), [pr#29094](#), Casey Bodley)
- rgw: radosgw-admin: fix syncs\_from in bucket sync status ([issue#40022](#), [pr#28243](#), Casey Bodley)
- rgw: radosgw-admin: sync status displays id of shard furthest behind ([pr#32311](#), Casey Bodley)
- rgw: radosgw-admin: update help for max-concurrent-ios ([pr#30742](#), Paul Emmerich)
- rgw: reduce per-shard entry count during ordered bucket listing ([pr#30853](#), J. Eric Ivancich)
- rgw: reject bucket tagging requests and document unsupported ([pr#26952](#), Casey Bodley)
- rgw: relax es zone validity check ([pr#32290](#), jiahui.zeng)
- rgw: release unused callback argument ([pr#32669](#), Ilsoo Byun)

- rgw: remove re-defined is\_tagging\_op in RGWHandler\_REST\_Bucket\_S3 ([pr#29004](#), zhang Shaowen)
- rgw: remove unused bucket parameter in check\_bucket\_shards ([pr#31186](#), zhang Shaowen)
- rgw: remove unused last\_run in reshards thread entry ([pr#31150](#), zhangshaowen)
- rgw: Replace COMPLETE\_MULTIPART\_MAX\_LEN with rgw\_max\_put\_param\_size ([issue#38002](#), [pr#26070](#), Lei Liu)
- rgw: replace direct calls to ioctx.operate() ([pr#28569](#), Ali Maredia)
- rgw: ReplaceKeyPrefixWith and ReplaceKeyWith can not set at the same time ([pr#32609](#), yuliyang)
- rgw: reshards list may return more than specified max\_entries ([pr#31355](#), zhangshaowen)
- rgw: rest client fixes for cloud sync XML outputs ([pr#27680](#), Abhishek Lekshmanan)
- rgw: return error if lock log shard fails ([pr#31344](#), zhangshaowen)
- rgw: return ERR\_NO SUCH BUCKET early while evaluating bucket policy ([issue#38420](#), [pr#26569](#), Abhishek Lekshmanan)
- rgw: rgw : Bucket mv, bucket chown and user rename utilities ([issue#35885](#), [issue#24348](#), [pr#28813](#), Shilpa Jagannath, Marcus Watts)
- rgw: rgw admin: add tenant argument to reshards cancel ([pr#26887](#), Abhishek Lekshmanan)
- rgw: rgw admin: disable stale instance delete in a multiste env ([pr#26852](#), Abhishek Lekshmanan)
- rgw: rgw multisite: add perf counters to data sync ([issue#38549](#), [pr#26722](#), Casey Bodley)
- rgw: rgw multisite: avoid writing bilog entries on PREPARE and CANCEL ([pr#26755](#), Casey Bodley)
- rgw: rgw multisite: data sync checks empty next\_marker for datalog ([issue#39033](#), [pr#27276](#), Casey Bodley)
- rgw: rgw multisite: enforce spawn window for incremental data sync ([pr#32534](#), Casey Bodley)
- rgw: rgw multisite: fixes for concurrent version creation ([pr#31325](#), Casey Bodley)
- rgw: rgw/kafka: add ssl+sasl security to kafka ([pr#31834](#), Yuval Lifshitz)

- rgw: rgw/multisite: Dont allow certain radosgw-admin commands to run on non-master zone ([issue#39548](#), [pr#28861](#), Shilpa Jagannath)
- rgw: rgw/multisite: warn if bucket chown command is run on non-master zone ([pr#32932](#), Shilpa Jagannath)
- rgw: rgw/multisite:RGWListBucketIndexesCR for data full sync pagination ([issue#39551](#), [pr#28146](#), Shilpa Jagannath)
- rgw: rgw/notification: add opaque data ([pr#32723](#), Yuval Lifshitz)
- rgw: rgw/pubsub: add kafka notification endpoint ([pr#30960](#), Yuval Lifshitz)
- rgw: rgw/pubsub: fix doc on updates. fix multi-notifications ([pr#27931](#), Yuval Lifshitz, Casey Bodley)
- rgw: rgw/pubsub: fix records/event json format to match documentation ([pr#31926](#), Yuval Lifshitz)
- rgw: rgw/pubsub: handle subscription conf errors better ([pr#27530](#), Yuval Lifshitz)
- rgw: rgw/pubsub: notification filtering by object tags ([pr#31878](#), Yuval Lifshitz)
- rgw: rgw/pubsub: prevent kafka thread from spinning when there are no messages ([pr#31998](#), Yuval Lifshitz)
- rgw: rgw/pubsub: send notifications from multi-delete op ([pr#32155](#), Yuval Lifshitz)
- rgw: rgw/pubsub: service reordering issue ([pr#29877](#), Yuval Lifshitz)
- rgw: rgw/rgw\_client\_io\_filters.h: print size\_t the portable way ([pr#28838](#), Kefu Chai)
- rgw: rgw/rgw\_crypt.cc: silence -Wsign-compare GCC warning ([pr#29151](#), Kefu Chai)
- rgw: rgw/rgw\_main: auto set radosgws cpu affinity according to numa\_node configuration ([pr#31001](#), luo rixin)
- rgw: rgw/rgw\_op: Remove get\_val from hotpath via legacy options ([pr#29943](#), Mark Nelson)
- rgw: rgw/rgw\_rados: set pg\_autoscale\_bias=4 for omap pools ([pr#27375](#), Sage Weil, Casey Bodley)
- rgw: rgw/rgw\_reshard: Dont dump RGWBucketReshard JSON in process\_single\_logshard ([pr#29894](#), Mark Nelson)
- rgw: rgw/rgw\_user: add [[maybe\_unused]] for silencing -Wunused-variable waxe2x80xa6 ([pr#30035](#), Kefu Chai)

- rgw: rgw/services: silence -Wunused-variable warning ([pr#30063](#), Lan Liu)
- rgw: RGW: add bucket permission verify when copy obj ([pr#29628](#), NancySu05)
- rgw: RGW: fix an endless loop error when to show usage ([pr#30470](#), lvshuhua)
- rgw: RGW: Set appropriate bucket quota value (when quota value is less than 0) ([pr#30920](#), GaryHyg)
- rgw: RGW: Listobjects v2 ([pr#28102](#), Albin Antony)
- rgw: RGWCoroutine::call(nullptr) sets retcode=0 ([pr#29856](#), Casey Bodley)
- rgw: rgwfile reqid: absorbs rgw\_file: allocate new id for continued request #25664 ([issue#37734](#), [pr#28108](#), Matt Benjamin, Tao Chen)
- rgw: RGWPeriodPusher uses zone system key for inter-zonegroup messages ([issue#39287](#), [pr#27576](#), Casey Bodley)
- rgw: RGWSI\_User\_Module filters .buckets objects out of user listing ([pr#29695](#), Casey Bodley)
- rgw: rgw\_file: advance\_mtime() should consider namespace expiration ([issue#40415](#), [pr#28632](#), Matt Benjamin)
- rgw: rgw\_file: all directories are virtual with respect to contents ([issue#40204](#), [pr#28451](#), Matt Benjamin)
- rgw: rgw\_file: avoid string::front() on empty path ([pr#32596](#), Matt Benjamin)
- rgw: rgw\_file: dont deadlock in advance\_mtime() ([pr#29560](#), Matt Benjamin)
- rgw: rgw\_file: fix readdir eof() calc-caller stop implies !eof ([issue#40375](#), [pr#28565](#), Matt Benjamin)
- rgw: rgw\_file: include tenant when hashing bucket names ([issue#40118](#), [pr#28370](#), Matt Benjamin)
- rgw: rgw\_file: introduce fast S3 Unix stats (immutable) ([issue#40456](#), [pr#28664](#), Matt Benjamin)
- rgw: rgw\_file: permit lookup\_handle to lookup root\_fh ([pr#28440](#), Matt Benjamin)
- rgw: rgw\_file: readdir: do not construct markers w/leading / ([pr#29670](#), Matt Benjamin)
- rgw: rgw\_file: save etag and acl info in setattr ([pr#26439](#), Tao Chen)
- rgw: rgw\_lc: use a new bl while encoding RGW\_ATTR\_LC ([pr#28049](#), Abhishek Lekshmanan)
- rgw: rgw\_sync: drop ENOENT error logs from mdlog ([pr#26908](#), Abhishek Lekshmanan)

- rgw: s/std::map/boost::container::flat\_map/ cls\_bucket\_list\_ordered ([pr#28637](#), Matt Benjamin)
- rgw: S3 compatible pubsub API ([pr#27091](#), Yuval Lifshitz)
- rgw: s3: dont require a body in S3 put-object-acl ([pr#31987](#), Matt Benjamin)
- rgw: save an unnecessary copy of RGWEnv ([pr#28426](#), Mark Kogan)
- rgw: Select the std::bitset to resolve ambiguity ([pr#31126](#), Willem Jan Withagen)
- rgw: set bucket attr twice when delete lifecycle config ([pr#30862](#), zhang Shaowen)
- rgw: set correct storage class for append ([pr#31088](#), yuliyang)
- rgw: set correct storage class for post object upload ([pr#30956](#), yuliyang)
- rgw: set null version object acl issues ([issue#36763](#), [pr#25044](#), Tianshan Qu)
- rgw: shard number must be non-negative when resharding the bucket ([pr#29037](#), zhang Shaowen)
- rgw: silence a -Wunused-function warning in pubsu ([pr#27578](#), Casey Bodley)
- rgw: Silence warning: control reaches end of non-void function ([issue#40747](#), [pr#28809](#), Jos Collin)
- rgw: split mdlog/datalog trimming into separate files ([pr#27579](#), Casey Bodley)
- rgw: sts: add all http args to req\_info ([pr#31661](#), yuliyang)
- rgw: support encoding-type param for list bucket multiparts ([pr#30993](#), Abhishek Lekshmanan)
- rgw: support radosgw-admin zone/zonegroup placement get command ([pr#30880](#), jiahuizeng)
- rgw: support specify user default placement and placement\_tags when create or modify user ([pr#31185](#), yuliyang)
- rgw: svc.bucket: assign to optional<> using = ([pr#32433](#), Kefu Chai)
- rgw: swift: bugfix: <https://tracker.ceph.com/issues/37765> ([pr#25962](#), Andrey Groshev)
- rgw: sync counters: drop spaces from counter names ([pr#27725](#), Abhishek Lekshmanan)
- rgw: sync with elastic search v7 ([pr#29637](#), Chang Liu)
- rgw: TempURL should not allow PUTs with the X-Object-Manifest ([issue#20797](#), [pr#16659](#), Radoslaw Zarzynski)

- rgw: test/rgw: fix test\_rgw\_reshard\_wait with -DHAVE\_BOOST\_CONTEXT=OFF ([pr#32811](#), Yaakov Selkowitz)
- rgw: test: modify iam tests to use a function to set bits ([pr#32808](#), Abhishek Lekshmanan)
- rgw: tests: Fix building with -DWITH\_BOOST\_CONTEXT=OFF ([pr#29430](#), Ulrich Weigand)
- rgw: the http response code of delete bucket should not be 204-no-content ([pr#30471](#), Chang Liu)
- rgw: Thread optional yield context through get\_bucket\_info call path ([pr#27898](#), Ali Maredia)
- rgw: thread option\_yield through bucket index transaction prepare ([pr#28152](#), Ali Maredia)
- rgw: unexpected crash when creating bucket in librgw ([pr#26089](#), Tao CHEN)
- rgw: update op\_mask of user via admin rest api ([issue#39084](#), [pr#21154](#), Ning Yao)
- rgw: update the hash source for multipart entries during resharding ([pr#32617](#), dongdong tao)
- rgw: update the radosgw-admin reshard status ([issue#37615](#), [pr#25496](#), Mark Kogan)
- rgw: updates to resharding documentation ([issue#39007](#), [pr#27250](#), J. Eric Ivancich)
- rgw: url decode PutUserPolicy params ([pr#29578](#), Abhishek Lekshmanan)
- rgw: url encode common prefixes for List Objects response ([pr#30970](#), Abhishek Lekshmanan)
- rgw: usage dump\_unsigned instead dump\_int ([pr#28308](#), yuliyang)
- rgw: usage dump\_unsigned instead dump\_int in dump\_usage\_categories\_info ([pr#25808](#), yuliyang)
- rgw: use bucket creation time from bucket instance info ([pr#32180](#), Yehuda Sadeh)
- rgw: use explicit to\_string() overload for boost::string\_ref ([issue#39611](#), [pr#28013](#), Casey Bodley)
- rgw: use new Stopped state for special handling of bucket sync disable ([pr#33054](#), Casey Bodley)
- rgw: use STSEngine::authenticate when post upload with x\_amz\_security\_token ([pr#31879](#), yuliyang)
- rgw: use the compatibility function for pthread\_setname ([pr#27456](#), Willem Jan Withagen)

- rgw: user policy: forward write requests to master zone ([pr#32476](#), Abhishek Lekshmanan)
- rgw: vstart: move [client.rgw] config into [client] ([pr#29778](#), Casey Bodley)
- rgw: vstart: only add -debug-ms=1 in RGWDEBUG ([pr#27409](#), Casey Bodley)
- rgw: warn on potential insecure mon connection ([pr#33777](#), Yehuda Sadeh)
- rgw: when resharding store progress json ([pr#30575](#), Mark Kogan)
- rgw: when you abort a multipart upload request, the quota may be not updated ([pr#29703](#), Richard Bai(xe7x99xbdx5xadxa6xe4xbdx99))
- rgw: Zipper - RGWUser ([pr#32298](#), Daniel Grynewicz)
- rgw: [RFC] rgw: raise default rgw\_bucket\_index\_max\_aio to 128 ([pr#28558](#), Casey Bodley)
- rgw: [rgw]:Validate bucket names as per revised s3 spec ([pr#26787](#), Soumya Koduri)
- seastar,crimson: pickup change to pin socket to fixed core ([pr#32797](#), Kefu Chai)
- seastar: pick up changes for better performance ([pr#28008](#), Kefu Chai)
- seastar: pick up latest changes and cleanups ([pr#29942](#), Kefu Chai)
- seastar: pick up the latest seastar ([pr#28709](#), Kefu Chai)
- seastar: pickup change to fix cgroups V2 support ([pr#32978](#), Kefu Chai)
- seastar: pickup the recent future optimizations ([pr#32296](#), Radoslaw Zarzynski)
- seastar: pickup unix domain socket support ([pr#30578](#), Kefu Chai)
- src/: silence GCC warnings ([pr#28684](#), Adam C. Emerson, Kefu Chai)
- src/msg/async/net\_handler.cc: Fix compilation ([pr#31637](#), Carlos Valiente)
- src/script/kubejacker: Fix and simplify ([issue#39065](#), [pr#27292](#), Sebastian Wagner)
- src/script: extract mypy config to mypy.ini ([pr#28264](#), Alfonso Martxc3xadnez)
- src/telemetry: remove, now lives in ceph-telemetry.git ([pr#31170](#), Dan Mick)
- src: polish the wording ([pr#33224](#), Jun Su)
- stop.sh: add -crimson option ([pr#28676](#), Kefu Chai)
- stop.sh: do not try to contact mon unless cluster is up ([pr#32295](#), Kefu Chai)
- support RDMA NIC without SRQ in msg/async/rdma ([pr#29947](#), Changcheng Liu, Roman Penyaev)

- tasks/ceph\_deploy: get rid of iteritems for python3 ([pr#30791](#), Kyr Shatskyy)
- telemetry: make server compensate for older mgr modules, elasticsearch ([pr#27802](#), Dan Mick)
- test/crimson: fix interpretability with perf\_async\_msgr ([pr#28913](#), Yingxin Cheng)
- tests,tools: ceph-objectstore-tool: call collection\_bits() crashes on the meta colxe2x80xa6 ([pr#31133](#), David Zafman)
- tests,tools: ceph-objectstore-tool: set log date format ([pr#29297](#), Robert Church)
- tests,tools: tools/ceph-dencoder: split types.h into smaller pieces ([issue#39595](#), [pr#28359](#), Kefu Chai)
- tests,tools: tools/setup-virtualenv.sh: do not default to python2.7 ([pr#30379](#), Nathan Cutler)
- tests: add missing header cmath to test/mon/test\_mon\_memory\_target.cc ([pr#30284](#), Su Yue)
- tests: ceph-object-corpus: pick up 15.0.0-539-g191ab33faf ([pr#27867](#), Kefu Chai)
- tests: cls/queue: add unit tests ([pr#33218](#), Yuval Lifshitz)
- tests: corrected issues with RBD tests under EL8 distros ([pr#32684](#), Jason Dillaman)
- tests: crimson/net: configure seastar to accept on a fixed core ([pr#32632](#), Yingxin Cheng)
- tests: crimson/test: add CBT based perf tests ([pr#29612](#), Kefu Chai)
- tests: crimson/test: v2 failover tests with crimson FailoverTestPeer ([pr#30162](#), Yingxin Cheng)
- tests: crush, test: update editor variables ([pr#30537](#), Kefu Chai)
- tests: fio\_ceph\_messenger: catch up v2 proto changes by using dummy auth ([pr#27264](#), Roman Penyaev)
- tests: import-generated.sh: use PATH to get ceph-dencoder ([pr#27573](#), Changcheng Liu)
- tests: introduce compiletest\_cxx11\_client for C++11 conformity ([pr#25395](#), Radoslaw Zarzynski)
- tests: lvm/deactivate: add unit tests, remove -all ([pr#32277](#), Jan Fajerski)
- tests: mgr/dashboard: ability to provide custom credentials for E2E tests ([pr#33549](#), Alfonso Martxc3xadnez)

- tests: mgr/dashboard: Add linter for unclosed HTML tags ([issue#40686](#), [pr#28916](#), Patrick Nawracay)
- tests: mgr/dashboard: add python-common to \$PYTHONPATH ([pr#29525](#), Kefu Chai)
- tests: mgr/dashboard: Added breadcrumb tests to Manager modules and Alerts menu ([pr#26853](#), Nathan Weinberg)
- tests: mgr/dashboard: Added breadcrumb tests to NFS menu ([pr#26850](#), Nathan Weinberg)
- tests: mgr/dashboard: Added breadcrumb tests to Object Gateway menu items ([pr#25451](#), Nathan Weinberg, Tiago Melo)
- tests: mgr/dashboard: comment failing QA suites out ([pr#30864](#), Tatjana Dehler)
- tests: mgr/dashboard: disable pylints -py3k flag ([pr#30078](#), Ernesto Puerta)
- tests: mgr/dashboard: E2E test to verify Configuration editing functionality ([pr#29216](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: Explicitly type page variables ([pr#29324](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: Fix e2e host test ([pr#30377](#), Tiago Melo)
- tests: mgr/dashboard: fix existing issues in user integration tests ([pr#30789](#), Tatjana Dehler)
- tests: mgr/dashboard: fix stray requests/error in Grafana unit test ([pr#33572](#), Patrick Seidensal)
- tests: mgr/dashboard: fix tasks.mgr.dashboard.test\_rgw suite ([pr#33426](#), Alfonso Martxc3xadnez)
- tests: mgr/dashboard: fix tests in order to match pg num conventions ([pr#31906](#), Tatjana Dehler)
- tests: mgr/dashboard: Improve e2e script ([pr#29101](#), Valentin Bajrami)
- tests: mgr/dashboard: RBD Image Purge Trash, Move to Trash and Restore ([pr#29673](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: reactivate dashboard test suites ([pr#32005](#), Tatjana Dehler)
- tests: mgr/dashboard: Reduce code duplication through TableActionComponent UnitTests ([issue#40399](#), [pr#28633](#), Patrick Nawracay)
- tests: mgr/dashboard: restore working directory after creating venv ([pr#32371](#), Kefu Chai)
- tests: mgr/dashboard: RGW bucket E2E Tests ([pr#28999](#), Adam King, Rafael Quintero)

- tests: mgr/dashboard: RGW user E2E Tests ([pr#29237](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: take portal\_ip\_addresses as a list ([pr#28495](#), Kefu Chai)
- tests: mgr/dashboard: Update formatting of e2e test files ([pr#29070](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: Updated existing E2E tests to match new format ([pr#27408](#), Nathan Weinberg)
- tests: mgr/dashboard: Verify fields on Configuration page ([pr#29583](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: Verify fields on OSDs page ([pr#29447](#), Adam King, Rafael Quintero)
- tests: mgr/dashboard: Wait for iSCSI target put and delete ([pr#30588](#), Ricardo Marques)
- tests: mgr/dashboard: Write E2E tests for pool creation, deletion and verification ([issue#40693](#), [issue#38093](#), [pr#28928](#), Patrick Nawracay)
- tests: mgr/orch: try harder when pickle fails to marshal an exception ([pr#33701](#), Kefu Chai)
- tests: mgr/ssh: add make check integration ([pr#31523](#), Sebastian Wagner)
- tests: mgr/tox: make run-tox.sh scripts more robust ([issue#39323](#), [pr#27614](#), Nathan Cutler)
- tests: osd-backfill-space.sh test failed in TEST\_backfill\_multi\_partial() ([issue#39333](#), [pr#27769](#), David Zafman)
- tests: pybind/mgr: apply\_drivegroups should return Sequence[Completion] ([pr#33977](#), Kefu Chai)
- tests: python: pin mypy requirement to mypy==0.770 ([pr#33926](#), Sebastian Wagner)
- tests: qa.tests: added smoke suite to the schedule on mimic,nautilus ([pr#28479](#), Yuri Weinstein)
- tests: qa/ceph-ansible: Disable dashboard ([pr#29916](#), Brad Hubbard)
- tests: qa/ceph-ansible: Move to ansible 2.8 ([issue#40602](#), [pr#28803](#), Brad Hubbard)
- tests: qa/ceph-ansible: Move to Nautilus ([pr#27013](#), Brad Hubbard)
- tests: qa/ceph-ansible: Replace pgs with pg\_num ([issue#40605](#), [pr#28807](#), Brad Hubbard)
- tests: qa/ceph-ansible: Upgrade ansible version ([pr#33379](#), Brad Hubbard)

- tests: qa/cephadm/smoke: run on opensuse\_15.1 ([pr#33338](#), Nathan Cutler)
- tests: qa/crontab/teuthology-cronjobs: fix suite-branch ([pr#27140](#), Neha Ojha)
- tests: qa/distros/all: add openSUSE 15.1, drop openSUSE 12.2 ([pr#30597](#), Nathan Cutler)
- tests: qa/distros: add SLE-12-SP3 and SLE-15-SP1 ([pr#31112](#), Nathan Cutler)
- tests: qa/orchestrator: do not test mon update 3 host1 ([pr#32023](#), Sage Weil, Kefu Chai)
- tests: qa/standalone/ceph-helpers: resurrect all OSD before waiting for health ([pr#28328](#), Kefu Chai)
- tests: qa/standalone/test\_ceph\_daemon: Fix ceph daemon standalone test ([pr#31440](#), Thomas Bechtold)
- tests: qa/suites/krbd: fsx with object-map and fast-diff ([pr#32376](#), Ilya Dryomov)
- tests: qa/suites/rados/cephadm/upgrade: add simple upgrade test ([pr#33343](#), Sage Weil)
- tests: qa/suites/rados/cephadm: deploy all monitoring components ([pr#33785](#), Sage Weil)
- tests: qa/suites/rados/perf/objectstore: do not symlink to qa/objectstore ([pr#30309](#), Neha Ojha)
- tests: qa/suites/rados/perf: test min recommended osd\_memory\_target ([pr#30347](#), Neha Ojha)
- tests: qa/suites/rados: whitelist POOL\_APP\_NOT\_ENABLED warning ([pr#29763](#), Kefu Chai)
- tests: qa/suites/upgrade/nutilus-x/parallel: restart mgr.x before mons ([pr#33705](#), Neha Ojha)
- tests: qa/suites/upgrade: use correct branch names ([pr#27764](#), Neha Ojha)
- tests: qa/suites: do not test luminous-x upgrade path ([pr#27112](#), Kefu Chai)
- tests: qa/tasks/cbt.py: add support for client\_endpoints ([pr#28522](#), Neha Ojha)
- tests: qa/tasks/cbt.py: change port to work with client\_endpoints ([pr#28442](#), Neha Ojha)
- tests: qa/tasks/cbt.py: use git -depth 1 for faster clone ([pr#29597](#), Kefu Chai)
- tests: qa/tasks/ceph.py: quote <kind> in command line ([pr#33775](#), Kefu Chai)
- tests: qa/tasks/ceph.py: remove unused variables ([pr#31005](#), Kefu Chai)

- tests: qa/tasks/ceph2: add support for shell, packaged ceph-daemon ([pr#31891](#), Sage Weil)
- tests: qa/tasks/cephfs\_test\_runner: setattr to class not instance ([pr#32571](#), Kefu Chai)
- tests: qa/tasks/ceph\_deploy: assume systemd and simplify shutdown wonkiness ([pr#29030](#), Sage Weil)
- tests: qa/tasks/ceph\_deploy: install python3.6 instead of python3.4 for py3 tests ([pr#27504](#), Kefu Chai)
- tests: qa/tasks/ceph\_manager.py: ignore errors in test\_pool\_min\_size ([issue#40533](#), [pr#28731](#), Kefu Chai)
- tests: qa/tasks/ceph\_manager: capture stderr for COT ([pr#33805](#), Kefu Chai)
- tests: qa/tasks/ceph\_manager: do not panic if pg\_num\_target is missing ([pr#30973](#), Kefu Chai)
- tests: qa/tasks/ceph\_manager: do not pick a pool if there is no pools ([pr#32519](#), Kefu Chai)
- tests: qa/tasks/mgr/dashboard/test\_health: add allow\_unknown in mgr\_map ([pr#30517](#), Kefu Chai)
- tests: qa/tasks/mgr/dashboard/test\_health: add missing field for test\_full\_health ([pr#29615](#), Kefu Chai)
- tests: qa/tasks/mgr/dashboard/test\_health: update schema ([pr#32122](#), Tatjana Dehler)
- tests: qa/tasks/mgr/dashboard/test\_mgr\_module: sync w/ telemetry ([pr#29461](#), Kefu Chai)
- tests: qa/tasks/mgr/dashboard: set pg\_num to 16 ([pr#32575](#), Kefu Chai)
- tests: qa/tasks/mgr/test\_orchestrator\_cli: fix mon update test ([pr#32428](#), Kefu Chai)
- tests: qa/tasks/mgr/test\_orchestrator\_cli: fix service action tests ([pr#32518](#), Kefu Chai)
- tests: qa/tasks/mgr/test\_orchestrator\_cli: fix test\_host\_ls ([pr#33477](#), Sage Weil)
- tests: qa/tasks/mgr/test\_progress.py: fix bug in 9b4dbf0 ([pr#29385](#), Kamoltat (Junior) Sirivadhna)
- tests: qa/tasks/mgr/test\_progress.py: s/ev/new\_event/ ([issue#40618](#), [pr#29368](#), Kefu Chai)
- tests: qa/tasks/mgr: set mgr module option with -force ([pr#32588](#), Kefu Chai)

- tests: qa/tasks/vstart\_runner: write string to StringIO ([pr#32438](#), Kefu Chai)
- tests: qa/tasks: call super classs setUp() ([pr#33325](#), Kefu Chai)
- tests: qa/tasks: py3 compat (tasks exercised by rados suites) ([pr#33709](#), Kyr Shatskyy, Kefu Chai)
- tests: qa/tasks: use items() for py3 compatibility ([pr#30813](#), Kyr Shatskyy)
- tests: qa/tests: filtered in only trusty ([issue#40195](#), [pr#28439](#), Yuri Weinstein)
- tests: qa/tests: added mimic-x on master run ([pr#29428](#), Yuri Weinstein)
- tests: qa/tests: added nautilus-p2p to cron ([pr#27218](#), Yuri Weinstein)
- tests: qa/tests: added nautilus-x run ([pr#27252](#), Yuri Weinstein)
- tests: qa/tests: added new client-upgrade-\\*-nautilus suites for jewel, luminous, mimic ([pr#28067](#), Yuri Weinstein)
- tests: qa/tests: added ragweed coverage to stress-split\\* upgrade suites ([issue#40467](#), [issue#40452](#), [pr#28931](#), Yuri Weinstein)
- tests: qa/tests: added ragweed coverage to stress-split\\* upgrade suites ([issue#40467](#), [issue#40452](#), [pr#28932](#), Yuri Weinstein)
- tests: qa/tests: added rgw into upgrade sequence to improve coverage ([pr#29406](#), Yuri Weinstein)
- tests: qa/tests: reduced distro to run to be random ([pr#28435](#), Yuri Weinstein)
- tests: qa/tests: reduced frequency for luminous and mimic runs ([pr#27057](#), Yuri Weinstein)
- tests: qa/tests: removed all runs for luminous - EOL ([pr#33186](#), Yuri Weinstein)
- tests: qa/tests: removed upgrade/client-upgrade-hammer becasue ubuntu 14.04 xe2x80xa6 ([pr#28518](#), Yuri Weinstein)
- tests: qa/tests: removed 1node and systemd tests as ceph-deploy is not actively developed ([issue#40207](#), [issue#40208](#), [pr#28455](#), Yuri Weinstein)
- tests: qa/valgrind.supp: generalize the whiterule for aes-128-gcm to help rgw suite ([issue#38827](#), [pr#28305](#), Radoslaw Zarzynski)
- tests: qa/workunits/cephadm/test\_cephadm: drop stray exit 0 ([pr#32622](#), Sage Weil)
- tests: qa/workunits/cephtool/test.sh: a handful fixes ([pr#31689](#), Kefu Chai)
- tests: qa/workunits/mon/config.sh: s|bin/ceph|ceph| ([pr#27147](#), Kefu Chai)
- tests: qa/workunits/rados/test\_crash.sh: do not rm coredump ([pr#32883](#), Kefu Chai)

- tests: qa/workunits/rados/test\_envlibrados\_for\_rocksdb: accomodate rocksdb cxe2x80xa6 ([pr#32143](#), Kefu Chai)
- tests: qa/workunits/rados/test\_envlibrados\_for\_rocksdb: install newer cmake ([pr#29584](#), Kefu Chai)
- tests: qa/workunits/rados/test\_librados\_build.sh: download from current branch ([pr#31693](#), Kefu Chai)
- tests: qa/workunits/rados/test\_librados\_build.sh: install build deps ([pr#28484](#), Kefu Chai)
- tests: qa/workunits/rest: Better detection of rest url ([pr#26604](#), Brad Hubbard)
- tests: qa: add .qa link ([pr#32363](#), Patrick Donnelly)
- tests: qa: Add basic mypy support for the qa directory ([pr#32495](#), Thomas Bechtold)
- tests: qa: add path to device output schema ([pr#32427](#), Kefu Chai)
- tests: qa: add RHEL 7.7 and use as RHEL7 default ([pr#29908](#), Patrick Donnelly)
- tests: qa: correct zap disk with ceph-deploy tool ([pr#31312](#), Changcheng Liu, Alfredo Deza)
- tests: qa: distro helper symlinks ([pr#28371](#), Patrick Donnelly)
- tests: qa: enable CRB repo for RHEL8 ([pr#32426](#), Kefu Chai)
- tests: qa: enable dashboard tests to be run with -suite rados/dashboard ([pr#30434](#), Nathan Cutler)
- tests: qa: Enable flake8 tox and fix failures ([pr#32129](#), Thomas Bechtold)
- tests: qa: fix all the fsx.sh-invoking yaml files to install dependencies ([pr#33959](#), Greg Farnum)
- tests: qa: fix lingering ceph-mgr-ssh -> ceph-mgr-cephadm refs ([pr#32250](#), Sage Weil)
- tests: qa: get rid of iterkeys for py3 compatibility ([pr#30873](#), Kyr Shatskyy)
- tests: qa: kernel.sh: update for read-only changes ([pr#31773](#), Ilya Dryomov)
- tests: qa: krbd\_exclusive\_option.sh: fixup for json.tool ordering change ([pr#32358](#), Ilya Dryomov)
- tests: qa: krbd\_exclusive\_option.sh: update for recent kernel changes ([pr#32088](#), Ilya Dryomov)
- tests: qa: rbd\_workunit\_suites\_fsx: install build dependencies ([pr#33412](#), Ilya

Dryomov)

- tests: qa: run cephadm/smoke on opensuse 15.2 instead of 15.1 ([pr#33535](#), Nathan Cutler)
- tests: qa: update krbd tests for python3 ([pr#31968](#), Ilya Dryomov)
- tests: qa: update krbd\_blkroset.t and add krbd\_get\_features.t ([pr#31771](#), Ilya Dryomov)
- tests: qa: whitelist FS\_DEGRADED ([pr#32549](#), Kefu Chai)
- tests: remove spurious whitespace ([pr#33848](#), Milind Changire)
- tests: Revert qa/tasks/cbt: include py2 deps on ubuntu for now ([pr#32512](#), Kefu Chai)
- tests: script/run-cbt.sh: add support for ceph-osd testing ([pr#30811](#), Radoslaw Zarzynski)
- tests: script/run-cbt.sh: always use python3 ([pr#30321](#), Kefu Chai)
- tests: script/run-cbt.sh: check option correctly ([pr#30287](#), Kefu Chai)
- tests: script/run-cbt.sh: set fs.aio-max-nr for seastar ([pr#31667](#), Kefu Chai)
- tests: script/run\_mypy: Support mypy 0.740 ([pr#31192](#), Sebastian Wagner)
- tests: script/run\_tox.sh: do not use python2 if we have python3 ([pr#31751](#), Kefu Chai)
- tests: selinux: Update the policy for RHEL8 ([pr#28290](#), Boris Ranto)
- tests: src/test, qa/suites/rados/thrash: add dedup test ([pr#28983](#), Myoungwon Oh)
- tests: src/test/compressor: Add missing gtest ([pr#33731](#), Willem Jan Withagen)
- tests: src/test: fix creating two different objects for testing chunked object ([issue#39282](#), [pr#27667](#), Myoungwon Oh)
- tests: src/valgrind.sup: replace with the teuthologys file. Whitelist OpenSSL ([pr#27265](#), Radoslaw Zarzynski)
- tests: tasks/ceph: drop testdir replacement in skeleton\_config ([pr#30829](#), Kyr Shatskyy)
- tests: tasks/ceph: get rid of iteritems for python3 ([pr#30792](#), Kyr Shatskyy)
- tests: test/bench\_log: add usage function ([pr#31723](#), Xuqiang Chen)
- tests: test/bufferlist.cc: encode/decode int64\_t instead of long ([pr#29881](#), Alexandre Oliva)

- tests: test/cli/ceph-conf: fix test ([pr#28818](#), Kefu Chai)
- tests: test/cli: Make the ceph-conf test more liberal ([pr#29405](#), Willem Jan Withagen)
- tests: test/common/test\_util: skip it if /etc/os-release does not exist ([pr#27927](#), Kefu Chai)
- tests: test/crimson/: use 256M mem and 1 cpu core for each test ([pr#29152](#), Kefu Chai)
- tests: test/crimson/perf\_async\_msgr: remove unused header file ([pr#28707](#), Jianpeng Ma)
- tests: test/crimson: add acceptable section to tests ([pr#30315](#), Kefu Chai)
- tests: test/crimson: add unit-test for ceph::net::Socket ([pr#28623](#), Yingxin Cheng)
- tests: test/crimson: cbt test does rand-reads instead of seq-reads ([pr#30794](#), Radoslaw Zarzynski)
- tests: test/crimson: fix a compiler error ([pr#27883](#), Jianpeng Ma)
- tests: test/crimson: fix build of unittest\_seastar\_monc ([pr#27515](#), Kefu Chai, Yingxin Cheng)
- tests: test/crimson: fix FTBFS ([pr#28902](#), Kefu Chai)
- tests: test/crimson: fix msgr test of ref counter racing ([issue#36405](#), [pr#28362](#), Yingxin Cheng)
- tests: test/crimson: implement a remote async TestPeer for crimson msgr tests ([pr#31156](#), Yingxin Cheng)
- tests: test/crimson: improved perf\_crimson\_msgr with timer and sampled lat ([pr#28542](#), Yingxin Cheng)
- tests: test/crimson: include writes in perf\_crimson/async\_server ([pr#27429](#), Yingxin Cheng)
- tests: test/crimson: lower the bar for cbt test ([pr#30458](#), Kefu Chai)
- tests: test/crimson: remove unittest\_seastar\_socket temporarily ([pr#32720](#), Kefu Chai)
- tests: test/crimson: update to accommodate Dispatcher changes ([pr#27093](#), Kefu Chai)
- tests: test/crimson: v2 failover tests with ack/keepalive ([pr#30803](#), Yingxin Cheng)

- tests: test/crimson: verify msgr v2 behavior with different policies ([pr#30925](#), Yingxin Cheng)
- tests: test/erasure-code: add exception handling to k & m ([pr#30087](#), Hang Li)
- tests: test/fio/fio\_ceph\_messenger: make exec multi client on the same host ([pr#28464](#), Jianpeng Ma)
- tests: test/fio: fix a compiler error ([pr#27880](#), Jianpeng Ma)
- tests: test/fio: introduce fio ioengine: fio\_ceph\_messenger ([pr#24678](#), Roman Penyaev)
- tests: test/kv\_store\_bench: Fix double free error ([pr#32439](#), Xuqiang Chen, luo rixin)
- tests: test/librados: avoid residual crush rule after test case execution ([issue#40970](#), [pr#29341](#), Bingyi Zhang)
- tests: test/librados: free AioCompletion using AioCompletion::release() ([pr#30204](#), Kefu Chai)
- tests: test/librados: use GTEST\_SKIP() to skip test ([pr#32770](#), Kefu Chai)
- tests: test/msgr: fix ComplexTest fail when using DPDK protocol stack ([pr#31910](#), Chunsong Feng)
- tests: test/msgr: make ceph\_perf\_msgr\_client/server work ([pr#28842](#), Jianpeng Ma)
- tests: test/objectstore: silence -Wsign-compare warning ([pr#27750](#), Kefu Chai)
- tests: test/old: remove stale tests ([pr#29124](#), Kefu Chai)
- tests: test/pybind/test\_ceph\_argparse.py: pg\_num of pool creation now optional ([pr#30535](#), xie xingguo)
- tests: test/python: remove stale tests ([pr#29413](#), Kefu Chai)
- tests: test/TestOSDScrub: fix mktime() error ([pr#33430](#), luo rixin)
- tests: test/test\_socket: fix dispatch\_sockets() unexpected exception ([pr#33482](#), luo rixin)
- tests: test/test\_weighted\_shuffle: enlarge epsilon ([pr#27181](#), Kefu Chai)
- tests: test/unittest\_bluefs: always remove temp bdev file ([pr#29676](#), Kefu Chai)
- tests: test/venv: do not hardwire to py2.7 for tox tests ([pr#29761](#), Willem Jan Withagen)
- tests: test: Add flush\_pg\_stats to avoid race with getting num\_shards\_repaired ([pr#33776](#), David Zafman)

- tests: test: Add #include <array> ([pr#27455](#), Willem Jan Withagen)
- tests: test: Allow fractional milliseconds to make test possible ([pr#30220](#), David Zafman)
- tests: test: do not include unnecessary includes ([pr#30065](#), Kefu Chai)
- tests: test: Do not test unicode if boost::spirit >= 1.72 ([pr#32388](#), Willem Jan Withagen)
- tests: test: Expect being off by up to 2 and make sure all PGs are active+clean ([pr#33566](#), David Zafman)
- tests: test: Fix failing ceph\_objectstore\_tool.py test ([pr#33593](#), David Zafman)
- tests: test: Fix race with osd restart and doing a scru ([pr#32039](#), David Zafman)
- tests: test: fix unused asserts variable in ceph\_test\_osd\_stale\_read.cc ([pr#32789](#), Radoslaw Zarzynski)
- tests: test: Fix wait\_for\_state() to wait for a PG to get into a state ([pr#32628](#), David Zafman)
- tests: test: Ignore OSD\_SLOW\_PING\_TIME\\* if injecting socket failures ([pr#30714](#), David Zafman)
- tests: test: move bluestore dependent code under WITH\_BLUESTORE ([pr#31335](#), Willem Jan Withagen)
- tests: test: remove Dockerfile for centos7 and add Dockerfile for centos8 ([pr#33452](#), Kefu Chai)
- tests: test: remove useless ASSERT\_XXX macros for rgw test ([pr#30062](#), Zhi Zhang)
- tests: test: silence warning unused variable nvme ([pr#33650](#), Jos Collin)
- tests: test: Update pg log test for new trimming behavior ([pr#32945](#), David Zafman)
- tests: use python3 compatible print ([pr#30758](#), Kyr Shatskyy)
- tests: vstart.sh: Make sure mkdir succeeds ([pr#30005](#), Willem Jan Withagen)
- test\_alien\_echo: update to use crimson:: namespace ([pr#31135](#), Samuel Just)
- test\_cephadm.sh: pass -fsid to shell command ([pr#32389](#), Sage Weil)
- test\_cephadm: use container shell for ceph cmds ([pr#32627](#), Michael Fritch)
- tools: add maxread in rados listomapkeys ([pr#30637](#), lvshuhua)
- tools: adding ceph level immutable obj cache daemon ([pr#25545](#), Yuan Zhou, Dehao)

Shang)

- tools: backport-create-issue: flush line before overprinting ([pr#31688](#), Nathan Cutler)
- tools: backport-create-issue: read redmine key from file ([pr#31533](#), Tiago Melo)
- tools: backport-create-issue: resolve parent if all backports resolved/rejected ([pr#30752](#), Nathan Cutler)
- tools: backport-create-issue: resolve parent only if parent has backport issues ([pr#31753](#), Nathan Cutler)
- tools: backport-resolve-issue: narrow regular expression and read key/token from files ([pr#31594](#), Nathan Cutler)
- tools: backport-resolve-issue: populate tracker\_description method ([pr#33105](#), Nathan Cutler)
- tools: backport-resolve-issue: recognize that Target version is populated and prune duplicate URLs ([pr#31247](#), Nathan Cutler)
- tools: backport-resolve-issue: resolve multiple backport issues ([pr#30988](#), Nathan Cutler)
- tools: backport-resolve-issue: use Basic Authentication instead of access\_token ([pr#33173](#), Nathan Cutler)
- tools: build-integration-branch: dont fail on existing branch ([pr#33093](#), Sage Weil)
- tools: build-integration-branch: take PRs in chronological order ([pr#31132](#), Nathan Cutler)
- tools: ceph-backport.sh: allow user to specify -fork explicitly ([pr#31734](#), Nathan Cutler)
- tools: ceph-backport.sh: automate setting of milestone and component label, implement -version option ([pr#30725](#), Nathan Cutler)
- tools: ceph-backport.sh: cherry-pick individual commits ([pr#30097](#), Jan Fajerski)
- tools: ceph-backport.sh: fix setup routine ([pr#33456](#), Nathan Cutler)
- tools: ceph-backport.sh: guess component with -existing-pr ([pr#31419](#), Nathan Cutler)
- tools: ceph-backport.sh: implement -milestones feature and more-careful vetting ([pr#30879](#), Nathan Cutler)
- tools: ceph-backport.sh: implement interactive setup routine and new options ([pr#31366](#), Nathan Cutler)

- tools: ceph-backport.sh: use Basic Authentication instead of access\_token ([pr#33182](#), Nathan Cutler)
- tools: ceph-conf: added -show-config-value to usage ([pr#29981](#), James McClune)
- tools: ceph-crash: use open(...,r) to read bytes for Python3 ([issue#40781](#), [pr#29053](#), Dan Mick)
- tools: ceph-daemon: ExecStart=/bin/bash script ([pr#31319](#), Sage Weil)
- tools: ceph-daemon: fix typo in the output\_pub\_ssh\_key argument ([pr#31337](#), John McGowan)
- tools: ceph-daemon: Fix ls cmd for legacy confs ([pr#31329](#), Michael Fritch)
- tools: ceph-monstore-tool: print out caps when rebuilding monstore ([pr#27340](#), Kefu Chai)
- tools: ceph-objectstore-tool: return 0 if incmap is sane ([pr#29704](#), Kefu Chai)
- tools: ceph-objectstore-tool: update-mon-db: do not fail if incmap is missing ([pr#29571](#), Kefu Chai)
- tools: ceph.in: fix verbose print ([pr#29486](#), luo.runbing)
- tools: cls: add timeindex types to ceph-dencoder ([pr#27780](#), Abhishek Lekshmanan)
- tools: github/codeowners: add ceph-volume ([pr#31883](#), Jan Fajerski)
- tools: github: Add CODEOWNERS for designated code-owner reviews ([pr#29451](#), Ernesto Puerta)
- tools: no-mon-config switch for ceph-objectstore-tool ([pr#26717](#), Igor Fedotov)
- tools: pin the version of breathe that works with Python2 ([pr#27721](#), Alfredo Deza)
- tools: script/backport-create-issue: add -resolve-parent feature ([pr#29904](#), Nathan Cutler)
- tools: script/backport-create-issue: handle long Redmine issue names ([pr#27887](#), Nathan Cutler)
- tools: script/backport-resolve-issue: better error message ([pr#30187](#), Nathan Cutler)
- tools: script/backport-resolve-issue: handle tracker URLs better ([pr#29950](#), Nathan Cutler)
- tools: script/ceph-backport-sh: add access\_token parameter to all ghub api cxe2x80xa6 ([pr#29261](#), Jan Fajerski)

- tools: script/ceph-backport.sh: Add prepare function ([pr#28446](#), Tiago Melo)
- tools: script/ceph-backport.sh: Allow to set component label ([pr#29318](#), Tiago Melo)
- tools: script/ceph-backport.sh: allow user to specify remote repo ([pr#27233](#), Kefu Chai)
- tools: script/ceph-backport.sh: carry https through to logical conclusion ([pr#29743](#), Nathan Cutler)
- tools: script/ceph-backport.sh: Fix verification of git repository ([pr#30398](#), Tiago Melo)
- tools: script/ceph-backport.sh: make the script idempotent ([pr#30106](#), Nathan Cutler)
- tools: script/ceph-backport.sh: Use secure access for tracker.ceph.com ([pr#29438](#), Willem Jan Withagen)
- tools: script/ceph-backport.sh: wholesale refactor ([pr#29957](#), Nathan Cutler)
- tools: script/ceph-release-notes: alternate merge commit format ([pr#27281](#), Nathan Cutler)
- tools: script/ptl-tool: update for python3 ([pr#29095](#), Patrick Donnelly)
- tools: script/run\_mypy: Sort groups ([pr#28225](#), Sebastian Wagner)
- tools: script/run\_tox.sh: remove unused code ([pr#30386](#), Kefu Chai)
- tools: script/sephia\_bt.sh: remove stale script ([pr#29129](#), Kefu Chai)
- tools: script: add backport-resolve-issue ([pr#29797](#), Nathan Cutler)
- tools: script: enable nautilus in backport scripts ([pr#26973](#), Nathan Cutler)
- tools: script: Obtain milestones via github API ([pr#27221](#), Lenz Grimmer)
- tools: script: raw\_input was renamed to input in py3 ([pr#30346](#), Patrick Donnelly)
- tools: scripts/kubejacker: Fix mgr\_plugins target for centos ([pr#28078](#), Sebastian Wagner)
- tools: scripts/run\_mypy: add .gitignore ([pr#27118](#), Sebastian Wagner)
- tools: scripts: use https url for redmine ([pr#29536](#), Patrick Donnelly)
- tools: src/script/backport-create-issue: implement -force option ([pr#30571](#), Nathan Cutler)
- tools: src/script/check\_commands.sh: fix grep regex class range ([pr#29161](#),

Valentin Bajrami)

- tools: src/script/unhexdump-C: script to reverse a hexdump -C style hexdump ([pr#29098](#), Sage Weil)
- tools: stop.sh: use bash shell to solve syntax error ([pr#32263](#), luo rixin)
- tools: tool/ceph-conf: s/global\_pre\_init()/global\_init()/[\(issue#7849, pr#29058](#), Kefu Chai)
- tools: tool: ceph\_monstore\_tool: -readable=0 => -readable ([pr#32265](#), simon gao)
- tools: tools/ceph-kvstore-tool: print db stats ([pr#27162](#), Igor Fedotov)
- tools: tools/osdmaptool.cc: do not use deprecated std::random\_shuffle() ([pr#31990](#), Kefu Chai)
- tools: tools/rados: update advisory lock break usage with -lock-cookie required ([pr#31348](#), Zhi Zhang)
- tools: vstart.sh: fix CEPH\_PORT check and cleanups ([pr#26782](#), Changcheng Liu, Kefu Chai)
- tools: vstart: add -inc-osd option ([pr#30512](#), xie xingguo)
- tools: vstart: add new option to pass list of block devices to bluestore ([pr#27518](#), Jeff Layton)
- tools: vstart: fix error when getting CMake variables with the same prefix ([pr#31962](#), Kiefer Chang)
- tools: vstart: fix run() invocation for rgw ([pr#28386](#), Casey Bodley)
- Update grafana dashboards ([issue#39652](#), [pr#28043](#), Jan Fajerski)
- vstart.sh: add an option to use crimson-osd ([pr#27108](#), chunmei Liu, Kefu Chai)
- vstart.sh: correct ceph-run path ([pr#27968](#), Changcheng Liu)
- vstart.sh: fix install of cephadm ssh keys from ~/.ssh ([pr#33647](#), Sage Weil)
- vstart.sh: Fix problem that all extra\_conf got merged into single line ([pr#28586](#), Adam Kupczyk)
- vstart.sh: move extra\_seastar\_args up in vstart.sh ([pr#32366](#), Chunmei Liu)
- vstart.sh: unify the indent ([pr#27995](#), Kefu Chai, Richael Zhuang)
- vstart\_runner: split unicode arguments into lists ([pr#28561](#), Rishabh Dave)

## v14.2.15 Nautilus

This is the 15th backport release in the Nautilus series. This release fixes a ceph-volume regression introduced in v14.2.13 and includes few other fixes. We recommend users to update to this release.

### Notable Changes

- ceph-volume: Fixes lvm batch -auto, which breaks backward compatibility when using non rotational devices only (SSD and/or NVMe).
- BlueStore: Fixes a bug in collection\_list\_legacy which makes pgs inconsistent during scrub when running mixed versions of osds, prior to 14.2.12 with newer.
- MGR: progress module can now be turned on/off, using the commands: `ceph progress on` and `ceph progress off`.

### Changelog

- ceph-volume: fix filestore/dmcrypt activate ([pr#38198](#), Guillaume Abrioux)
- ceph-volume: fix lvm batch auto with full SSDs ([pr#38046](#), Dimitri Savineau, Guillaume Abrioux)
- os/bluestore: fix “end reached” check in collection\_list\_legacy ([pr#38100](#), Mykola Golub)
- mgr/progress: introduce turn off/on feature ([pr#38173](#), kamoltat)

## v14.2.14 Nautilus

This is the 14th backport release in the Nautilus series. This release fixes a security flaw affecting Messenger v2, among other fixes across components. We recommend users to update to this release.

### Notable Changes

- CVE 2020-25660: CEPHX\_V2 replay attack protection lost, for Messenger v2 (Ilya Dryomov)

### Changelog

- mgr/dashboard: Strange iSCSI discovery auth behavior ([pr#37333](#), Volker Theile)

- mgr/dashboard: redirect to original URL after successful login ([pr#36834](#), Avan Thakkar)
- mgr/prometheus: add pool compression stats ([pr#37563](#), Paul Cuzner)
- bluestore: test/objectstore/store\_test: kill ExcessiveFragmentation test case ([pr#37824](#), Igor Fedotov)
- bluestore: BlockDevice.cc: use pending\_aios instead of iovec size as ios num ([pr#37823](#), weixinwei)
- bluestore: Support flock retry ([pr#37842](#), Kefu Chai, wanghongxu)
- bluestore: attach csum for compressed blobs ([pr#37843](#), Igor Fedotov)
- osdc/ObjectCacher: overwrite might cause stray read request callbacks ([pr#37813](#), Jason Dillaman)
- mgr: avoid false alarm of MGR\_MODULE\_ERROR ([pr#38069](#), Kefu Chai, Sage Weil)
- mgr: fix race between module load and notify ([pr#37844](#), Mykola Golub, Patrick Donnelly)
- mon: set session\_timeout when adding to session\_map ([pr#37554](#), Ilya Dryomov)
- mon/MonClient: bring back CEPHX\_V2 authorizer challenges (Ilya Dryomov)
- osd/osd-rep-recov-eio.sh: TEST\_rados\_repair\_warning: return 1 ([pr#37815](#), David Zafman)
- rbd: librbd: ignore -ENOENT error when disabling object-map ([pr#37814](#), Jason Dillaman)
- rbd: rbd-nbd: don't ignore namespace when unmapping by image spec ([pr#37811](#), Mykola Golub)
- rgw/rgw\_file: Fix the incorrect lru object eviction ([pr#37804](#), luo rixin)
- rgw: fix expiration header returned even if there is only one tag in the object the same as the rule ([pr#37806](#), Or Friedmann)
- rgw: fix: S3 API KeyCount incorrect return ([pr#37810](#), 胡玮文)
- rgw: radosgw-admin should paginate internally when listing bucket ([pr#37802](#), J. Eric Ivancich)
- rgw: rgw\_file: avoid long-ish delay on shutdown ([pr#37552](#), Matt Benjamin)
- rgw: use yum rather than dnf for teuthology testing of rgw-orphan-list ([pr#37805](#), J. Eric Ivancich)

## v14.2.13 Nautilus

This is the 13th backport release in the Nautilus series. This release fixes a regression introduced in v14.2.12, and a few ceph-volume & RGW fixes. We recommend users to update to this release.

## Notable Changes

- Fixed a regression that caused breakage in clusters that referred to ceph-mon hosts using dns names instead of ip addresses in the `mon_host` param in `ceph.conf` ([issue#47951](#))
- ceph-volume: the `lvm batch` subcommand received a major rewrite

## Changelog

- ceph-volume: major batch refactor ([pr#37522](#), Jan Fajerski)
- mgr/dashboard: Proper format iSCSI target portals ([pr#37060](#), Volker Theile)
- rpm: move python-enum34 into rhel 7 conditional ([pr#37747](#), Nathan Cutler)
- mon/MonMap: fix unconditional failure for `init_with_hosts` ([pr#37816](#), Nathan Cutler, Patrick Donnelly)
- rgw: allow rgw-orphan-list to note when rados objects are in namespace ([pr#37799](#), J. Eric Ivancich)
- rgw: fix setting of namespace in ordered and unordered bucket listing ([pr#37798](#), J. Eric Ivancich)

## v14.2.12 Nautilus

This is the 12th backport release in the Nautilus series. This release brings a number of bugfixes across all major components of Ceph. We recommend that all Nautilus users upgrade to this release.

## Notable Changes

- The `ceph df` command now lists the number of pgs in each pool.
- Monitors now have a config option `mon_osd_warn_num_repaired`, 10 by default. If any OSD has repaired more than this many I/O errors in stored data a `OSD_TOO_MANY_REPAIRS` health warning is generated. In order to allow clearing of the warning, a new command `ceph tell osd.# clear_shards_repaired [count]` has been added. By default it will

set the repair count to 0. If you wanted to be warned again if additional repairs are performed you can provide a value to the command and specify the value of `mon_osd_warn_num_repaired`. This command will be replaced in future releases by the health mute/unmute feature.

- It is now possible to specify the initial monitor to contact for Ceph tools and daemons using the `mon_host_override` config option or `--mon-host-override <ip>` command-line switch. This generally should only be used for debugging and only affects initial communication with Ceph's monitor cluster.
- Fix an issue with osdmmaps not being trimmed in a healthy cluster ([issue#47296](#), [pr#36982](#))

## Changelog

---

- bluestore/bluefs: make accounting resiliant to unlock() ([pr#36909](#), Adam Kupczyk)
- bluestore: Rescue procedure for extremely large bluefs log ([pr#36930](#), Adam Kupczyk)
- bluestore: dump onode that has too many spanning blobs ([pr#36756](#), Igor Fedotov)
- bluestore: enable more flexible bluefs space management by default ([pr#37091](#), Igor Fedotov)
- bluestore: fix collection\_list ordering ([pr#37051](#), Mykola Golub)
- ceph-iscsi: selinux fixes ([pr#36304](#), Mike Christie)
- ceph-volume: add tests for new functions that run LVM commands ([pr#36615](#), Rishabh Dave)
- ceph-volume: dont use container classes in api/lvm.py ([pr#35878](#), Guillaume Abrioux, Rishabh Dave')
- ceph-volume: fix journal size argument not work ([pr#37377](#), wanghongxu)
- ceph-volume: fix simple activate when legacy osd ([pr#37195](#), Guillaume Abrioux)
- ceph-volume: fix test\_lvm.TestVolume.test\_is\_not\_ceph\_device ([pr#36493](#), Jan Fajerski)
- ceph-volume: handle idempotency with batch and explicit scenarios ([pr#35881](#), Andrew Schoen)
- ceph-volume: remove container classes from api/lvm.py ([pr#36610](#), Rishabh Dave)
- ceph-volume: remove unneeded call to get\_devices() ([pr#37413](#), Marc Gariepy)
- ceph-volume: report correct rejected reason in inventory if device type is

- invalid ([pr#36453](#), Satoru Takeuchi)
- ceph-volume: retry when acquiring lock fails ([pr#36926](#), Sxc3xa9bastien Han)
- ceph-volume: simple scan should ignore tmpfs ([pr#36952](#), Andrew Schoen)
- ceph.in: ignore failures to flush stdout ([pr#37226](#), Dan van der Ster)
- ceph.spec.in, debian/control: add smartmontools and nvme-cli dependencies ([pr#37288](#), Yaarit Hatuka)
- cephfs-journal-tool: fix incorrect read\_offset when finding missing objects ([pr#37479](#), Xue Yantao)
- cephfs: client: fix extra open ref decrease ([pr#36966](#), Xiubo Li)
- cephfs: client: make Client::open() pass proper cap mask to path\_walk ([pr#37231](#), "Yan, Zheng")
- cephfs: mds/CINode: Optimize only pinned by subtrees check ([pr#36965](#), Mark Nelson)
- cephfs: mds: After restarting an mds, its standy-replay mds remained in the "resolve" state ([pr#37179](#), Wei Qiaomiao)
- cephfs: mds: do not defer incoming mgrmap when mds is laggy ([issue#44638](#), [pr#36168](#), Nathan Cutler, Venky Shankar)
- cephfs: mds: fix incorrect check for if dirfrag is being fragmented ([pr#37035](#), "Yan, Zheng")
- cephfs: mds: fix mds forwarding request no\_available\_op\_found ([pr#36963](#), Yanhu Cao')
- cephfs: mds: fix purge\_queues \_calculate\_ops is inaccurate ([pr#37481](#), Yanhu Cao')
- cephfs: mds: kcephfs parse dirfrags ndist is always 0 ([pr#37177](#), Yanhu Cao')
- cephfs: mds: place MDSGatherBuilder on the stack ([pr#36967](#), Patrick Donnelly)
- cephfs: mds: recover files after normal session close ([pr#37178](#), "Yan, Zheng")
- cephfs: mds: resolve SIGSEGV in waiting for uncommitted fragments ([pr#36968](#), Patrick Donnelly)
- cephfs: osdc/Journaler: do not call onsafe->complete() if onsafe is 0 ([pr#37229](#), Xiubo Li)
- client: handle readdir reply without Fs cap ([pr#37232](#), "Yan, Zheng")
- common, osd: add sanity checks around osd\_scrub\_max\_premptions ([pr#37470](#), xie xingguo)

- common/config: less noise about configs from mon we cant apply ([pr#36289](#), Sage Weil')
- common: ignore SIGHUP prior to fork ([issue#46269](#), [pr#36181](#), Willem Jan Withagen, hzwuhongsong)
- compressor: Add a config option to specify Zstd compression level ([pr#37254](#), Bryan Stillwell)
- core: include/encoding: Fix encode/decode of float types on big-endian systems ([pr#37033](#), Ulrich Weigand)
- doc/rados: Fix osd\_op\_queue default value ([pr#36354](#), Benoxc3xaet Knecht)
- doc/rados: Fix osd\_scrub\_during\_recovery default value ([pr#37472](#), Benoxc3xaet Knecht)
- doc/rbd: add rbd-target-gw enable and start ([pr#36415](#), Zac Dover)
- doc: enable Read the Docs ([pr#37204](#), Kefu Chai)
- krbd: optionally skip waiting for udev events ([pr#37284](#), Ilya Dryomov)
- kv/RocksDBStore: make options compaction\_threads/disableWAL/flusher\_txe2x80xa6 ([pr#37055](#), Jianpeng Ma)
- librados: add LIBRADOS\_SUPPORTS\_GETADDRS support ([pr#36853](#), Xiubo Li, Jason Dillaman, Kaleb S. KEITHLEY, Kefu Chai)
- messages,mds: Fix decoding of enum types on big-endian systems ([pr#36814](#), Ulrich Weigand)
- mgr/balancer: use “==” and “!=” for comparing str ([pr#37471](#), Kefu Chai)
- mgr/dashboard/api: increase API health timeout ([pr#36607](#), Ernesto Puerta)
- mgr/dashboard: Allow editing iSCSI targets with initiators logged-in ([pr#37278](#), Tiago Melo)
- mgr/dashboard: Disabling the form inputs for the read\_only modals ([pr#37241](#), Nizamudeen)
- mgr/dashboard: Dont use any xlf file when building the default language ([pr#37550](#), Sebastian Krah')
- mgr/dashboard: Fix many-to-many issue in host-details Grafana dashboard ([pr#37306](#), Patrick Seidensal)
- mgr/dashboard: Fix pool renaming functionality ([pr#37510](#), Stephan Mxc3xbcller, Ernesto Puerta)
- mgr/dashboard: Hide table action input field if limit=0 ([pr#36783](#), Volker Theile)

- mgr/dashboard: Monitoring: Fix for the infinite loading bar action ([pr#37161](#), Nizamudeen A)
- mgr/dashboard: REST API returns 500 when no Content-Type is specified ([pr#37307](#), Avan Thakkar)
- mgr/dashboard: Unable to edit iSCSI logged-in client ([pr#36613](#), Ricardo Marques)
- mgr/dashboard: cpu stats incorrectly displayed ([pr#37295](#), Avan Thakkar)
- mgr/dashboard: document Prometheus security model ([pr#36920](#), Patrick Seidensal)
- mgr/dashboard: fix broken backporting ([pr#37505](#), Ernesto Puerta)
- mgr/dashboard: fix perf. issue when listing large amounts of buckets ([pr#37280](#), Alfonso Martxc3xadnez)
- mgr/dashboard: fix pool usage calculation ([pr#37309](#), Ernesto Puerta)
- mgr/dashboard: remove “This week/month/year” and “Today” time stamps ([pr#36790](#), Avan Thakkar)
- mgr/dashboard: table detail rows overflow ([pr#37324](#), Aashish Sharma)
- mgr/dashboard: wait longer for health status to be cleared ([pr#36784](#), Tatjana Dehler)
- mgr/devicehealth: fix daemon filtering before scraping device ([pr#36741](#), Yaarit Hatuka)
- mgr/diskprediction\_local: Fix array size error ([pr#36578](#), Benoxc3xaet Knecht)
- mgr/prometheus: automatically discover RBD pools for stats gathering ([pr#36412](#), Jason Dillaman)
- mgr/restful: use dict.items() for py3 compatible ([pr#36670](#), Kefu Chai)
- mgr/status: metadata is fetched async ([pr#37558](#), Michael Fritch)
- mgr/telemetry: fix device id splitting when anonymizing serial ([pr#37318](#), Yaarit Hatuka)
- mgr/volumes: add global lock debug ([pr#36828](#), Patrick Donnelly)
- mgr: Add missing states to PG\_STATES in mgr\_module.py ([pr#36785](#), Harley Gorrell)
- mgr: decrease pool stats if pg was removed ([pr#37476](#), Aleksei Gutikov)
- mgr: don't update pending service map epoch on receiving map from mon ([pr#37181](#), Mykola Golub')
- minor tweaks to fix compile issues under latest Fedora ([pr#36726](#), Willem Jan

Withagen, Kaleb S. KEITHLEY, Kefu Chai)

- mon/OSDMonitor: only take in osd into consideration when trimming osdmaps ([pr#36982](#), Kefu Chai)
- mon/PGMap: add pg count for pools in the ceph df command ([pr#36944](#), Vikhyat Umrao)
- mon: Warn when too many reads are repaired on an OSD ([pr#36379](#), David Zafman)
- mon: fix the Error ERANGE message when conf "osd\_objectstore" is filestore' ([pr#37474](#), wangyunqing')
- mon: mark pgtemp messages as no\_reply more consistently in preprocess\\_xe2x80xa6 ([pr#37171](#), Greg Farnum)
- mon: store mon updates in ceph context for future MonMap instantiation ([pr#36704](#), Patrick Donnelly, Shyamsundar Ranganathan)
- monclient: schedule first tick using mon\_client\_hunt\_interval ([pr#36634](#), Mykola Golub)
- msg/async/ProtocolV2: allow rxbuf/txbuf get bigger in testing ([pr#37081](#), Ilya Dryomov)
- osd/OSDCap: rbd profile permits use of "rbd\_info" ([pr#36413](#), Florian Florensa)
- osd/PeeringState: prevent peers num\_objects going negative ([pr#37473](#), xie xingguo')
- prometheus: Properly split the port off IPv6 addresses ([pr#36984](#), Matthew Oliver)
- rbd: include RADOS namespace in krbd symlinks ([pr#37468](#), Ilya Dryomov)
- rbd: librbd: Align rbd\_write\_zeroes declarations ([pr#36712](#), Corey Bryant)
- rbd: librbd: dont resend async\_complete if watcher is unregistered ([pr#37040](#), Mykola Golub')
- rbd: librbd: global and pool-level config overrides require image refresh to apply ([pr#36725](#), Jason Dillaman)
- rbd: librbd: using migration abort can result in the loss of data ([pr#37165](#), Jason Dillaman)
- rbd: make common options override krbd-specific options ([pr#37407](#), Ilya Dryomov)
- rgw/cls: preserve olh entrys name on last unlink ([pr#37462](#), Casey Bodley')
- rgw: Add bucket name to bucket stats error logging ([pr#37378](#), Seena Fallah)
- rgw: Empty reqs\_change\_state queue before unregistered\_reqs ([pr#37461](#), Soumya

Koduri)

- rgw: Expiration days cant be zero and transition days can be zero ([pr#37465](#), zhang Shaowen')
- rgw: RGWObjVersionTracker tracks version over increments ([pr#37459](#), Casey Bodley)
- rgw: Swift API anonymous access should 401 ([pr#37438](#), Matthew Oliver)
- rgw: add access log to the beast frontend ([pr#36727](#), Mark Kogan)
- rgw: add negative cache to the system object ([pr#37460](#), Or Friedmann)
- rgw: append obj: prevent tail from being GCed ([pr#36390](#), Abhishek Lekshmanan')
- rgw: dump transitions in RGWLifecycleConfiguration::dump() ([pr#36880](#), Shengming Zhang)
- rgw: fail when get/set-bucket-versioning attempted on a non-existent xe2x80xa6 ([pr#36188](#), Matt Benjamin)
- rgw: fix boost::asio::async\_write() does not return error ([pr#37157](#), Mark Kogan)
- rgw: fix double slash (//) killing the gateway ([pr#36682](#), Theofilos Mouratidis)
- rgw: fix shutdown crash in RGWAsyncReadMDLogEntries ([pr#37463](#), Casey Bodley)
- rgw: hold reloader using unique\_ptr ([pr#36770](#), Kefu Chai)
- rgw: log resharding events at level 1 (formerly 20) ([pr#36843](#), Or Friedmann)
- rgw: ordered bucket listing code clean-up ([pr#37169](#), J. Eric Ivancich)
- rgw: policy: reuse eval\_principal to evaluate the policy principal ([pr#36637](#), Abhishek Lekshmanan)
- rgw: radosgw-admin: period pull command is not always a raw\_storage\_op ([pr#37464](#), Casey Bodley)
- rgw: replace +with "%20" in canonical query string for s3 v4 auth' ([pr#37467](#), yuliyang\_yewu')
- rgw: urlencode bucket name when forwarding request ([pr#37435](#), caolei)
- run-make-check.sh: extract run-make.sh + run sudo with absolute path ([pr#36494](#), Kefu Chai, Ernesto Puerta)
- systemd: Support Graceful Reboot for AIO Node ([pr#37301](#), Wong Hoi Sing Edison)
- tools/osdmaptool.cc: add ability to clean\_temps ([pr#37477](#), Neha Ojha)
- tools/rados: Set locator key when exporting or importing a pool ([pr#37475](#), Iain Buclaw)

## v14.2.11 Nautilus

This is the eleventh backport release in the Nautilus series. This release brings a number of bugfixes across all major components of Ceph. We recommend that all Nautilus users upgrade to this release.

## Notable Changes

- RGW: The `radosgw-admin` sub-commands dealing with orphans – `radosgw-admin orphans find`, `radosgw-admin orphans finish`, `radosgw-admin orphans list-jobs` – have been deprecated. They have not been actively maintained and they store intermediate results on the cluster, which could fill a nearly-full cluster. They have been replaced by a tool, currently considered experimental, `rgw-orphan-list`.
- Now when noscrub and/or nodeep-scrub flags are set globally or per pool, scheduled scrubs of the type disabled will be aborted. All user initiated scrubs are NOT interrupted.
- Fixed a ceph-osd crash in `_committed_osd_maps` when there is a failure to encode the first incremental map. [issue#46443](#)

## Changelog

- bluestore: core: os/bluestore: fix large (>2GB) writes when bluefs\_buffered\_io = true ([pr#35404](#), Igor Fedotov)
- bluestore: os/bluestore: implement Hybrid allocator ([pr#35500](#), Adam Kupczyk, Kefu Chai, Igor Fedotov, xie xingguo)
- build/ops: build/ops: selinux: allow ceph\_t amqp\_port\_t:tcp\_socket ([pr#36190](#), Kaleb S. KEITHLEY, Thomas Serlin)
- ceph-volume: add dmcrypt support in raw mode ([pr#35831](#), Guillaume Abrioux)
- cephfs,pybind: pybind/cephfs: fix custom exception raised by cephfs.pyx ([pr#36180](#), Ramana Raja)
- cephfs: ceph\_fuse: add the '-d' option back for libfuse ([pr#35398](#), Xiubo Li)
- cephfs: client: fix directory inode can not call release callback ([pr#36177](#), sephia-liu)
- cephfs: client: fix setxattr for 0 size value (NULL value) ([pr#36173](#), Sidharth Anupkrishnan)
- cephfs: client: fix snap directory atime ([pr#36169](#), Luis Henriques)
- cephfs: client: introduce timeout for client shutdown ([issue#44276](#), [pr#36215](#),

Venky Shankar)

- cephfs: client: release the client\_lock before copying data in read ([pr#36294](#), Chencan)
- cephfs: client: static dirent for readdir is not thread-safe ([pr#36511](#), Patrick Donnelly)
- cephfs: mds: add config to require forward to auth MDS ([pr#35377](#), simon gao)
- cephfs: mds: cleanup uncommitted fragments before mds goes to active ([pr#35397](#), "Yan, Zheng")
- cephfs: mds: do not raise "client failing to respond to cap release" when client working set is reasonable ([pr#36513](#), Patrick Donnelly)
- cephfs: mds: do not submit omap\_rm\_keys if the dir is the basedir of merge ([pr#36178](#), Chencan)
- cephfs: mds: fix filelock state when Fc is issued ([pr#35841](#), Xiubo Li)
- cephfs: mds: fix hang issue when accessing a file under a lost parent directory ([pr#36179](#), Zhi Zhang)
- cephfs: mds: fix nullptr dereference in MDCache::finish\_rollback ([pr#36439](#), "Yan, Zheng")
- cephfs: mds: flag backtrace scrub failures for new files as okay ([pr#35400](#), Milind Changire)
- cephfs: mds: initialize MDSSlaveUpdate::waiter ([pr#36462](#), "Yan, Zheng")
- cephfs: mds: make threshold for MDS\_TRIM configurable ([pr#36175](#), Paul Emmerich)
- cephfs: mds: preserve ESlaveUpdate logevent until receiving OP\_FINISH ([pr#35394](#), Varsha Rao, songxinying)
- cephfs: mds: reset heartbeat in EMetaBlob replay ([pr#36170](#), Yanhu Cao)
- cephfs: mgr/fs/volumes misc fixes ([pr#36167](#), Patrick Donnelly, Kotresh HR, Ramana Raja)
- cephfs: mgr/volumes: Add snapshot info command ([pr#35672](#), Kotresh HR)
- cephfs: mgr/volumes: Deprecate protect/unprotect CLI calls for subvolume snapshots ([pr#36166](#), Shyamsundar Ranganathan)
- cephfs: qa: add debugging for volumes plugin use of libcephfs ([pr#36512](#), Patrick Donnelly)
- cephfs: qa: skip cache\_size check ([pr#36526](#), Patrick Donnelly)

- cephfs: tools/cephfs: don't bind to public\_addr ([pr#35401](#), "Yan, Zheng")
- cephfs: vstart\_runner: set mounted to True at the end of mount() ([pr#35396](#), Rishabh Dave)
- core,mon: mon/OSDMonitor: Reset grace period if failure interval exceeds a threshold ([pr#35798](#), Sridhar Seshasayee)
- core: mgr/DaemonServer.cc: make 'config show' on fsid work ([pr#36074](#), Neha Ojha)
- core: mgr/alert: can't set inventory\_cache\_timeout/service\_cache\_timeout from CLI ([pr#36104](#), Kiefer Chang)
- core: osd/PG: fix history.same\_interval\_since of merge target again ([pr#36161](#), xie xingguo)
- core: osd/PeeringState.h: Fix pg stuck in WaitActingChange ([pr#35389](#), chen qizhang)
- core: osd: Cancel in-progress scrubs (not user requested) ([pr#36292](#), David Zafman)
- core: osd: fix crash in \_committed\_osd\_maps if incremental osdmap crc fails ([pr#36339](#), Neha Ojha, Dan van der Ster)
- core: osd: make "missing incremental map" a debug log message ([pr#35386](#), Nathan Cutler)
- core: osd: make message cap option usable again ([pr#35738](#), Neha Ojha, Josh Durgin)
- mgr/dashboard: Allow to edit iSCSI target with active session ([pr#35998](#), Ricardo Marques)
- mgr/dashboard: Prevent dashboard breakdown on bad pool selection ([pr#35367](#), Stephan Müller)
- mgr/dashboard: Prometheus query error in the metrics of Pools, OSDs and RBD images ([pr#35884](#), Avan Thakkar)
- mgr/dashboard: add popover list of Stand-by Managers & Metadata Servers (MDS) in landing page ([pr#34095](#), Kiefer Chang, Avan Thakkar)
- mgr/dashboard: fix Source column i18n issue in RBD configuration tables ([pr#35822](#), Kiefer Chang)
- mgr/k8sevents: sanitise kubernetes events ([pr#35563](#), Paul Cuzner)
- mgr/prometheus: improve Prometheus module cache ([pr#35918](#), Patrick Seidensal)
- mgr: mgr/progress: Skip pg\_summary update if \_events dict is empty ([pr#36075](#), Manuel Lausch)

- mgr: mgr/telemetry: force -license when sending while opted-out ([pr#35390](#), Yaarit Hatuka)
- mgr: mon/PGMap: do not consider changing pg stuck ([pr#35959](#), Kefu Chai)
- monitoring: fixing some issues in RBD detail dashboard ([pr#35464](#), Kiefer Chang)
- msgr: New msgr2 crc and secure modes (msgr2.1) ([pr#35733](#), Jianpeng Ma, Ilya Dryomov)
- rbd: librbd: new 'write\_zeroes' API methods to supplement the discard APIs ([pr#36250](#), Jason Dillaman)
- rbd: mgr/dashboard: work with v1 RBD images ([pr#35712](#), Ernesto Puerta)
- rbd: rbd: librbd: Watcher should not attempt to re-watch after detecting blacklisting ([pr#35385](#), Jason Dillaman)
- rgw,tests: test/rgw: update hadoop versions ([pr#35778](#), Casey Bodley, Vasu Kulkarni)
- rgw: Add subuser to OPA request ([pr#36187](#), Seena Fallah)
- rgw: Add support wildcard subuser for bucket policy ([pr#36186](#), Seena Fallah)
- rgw: add "rgw-orphan-list" tool and "radosgw-admin bucket radoslist ..." ([pr#34127](#), J. Eric Ivancich)
- rgw: add check for index entry's existing when adding bucket stats during bucket reshards ([pr#36189](#), zhang Shaowen)
- rgw: add quota enforcement to CopyObj ([pr#36184](#), Casey Bodley)
- rgw: bucket list/stats truncates for user w/ >1000 buckets ([pr#36165](#), J. Eric Ivancich)
- rgw: cls\_bucket\_list\_(un)ordered should clear results collection ([pr#36163](#), J. Eric Ivancich)
- rgw: fix loop problem with swift stat on account ([pr#36185](#), Marcus Watts)
- rgw: lc: fix Segmentation Fault when the tag of the object was not found ([pr#36086](#), yupeng chen, zhuo li)
- rgw: ordered listing lcv not managed correctly ([pr#35882](#), J. Eric Ivancich)
- rgw: radoslist incomplete multipart uploads fix marker progression ([pr#36191](#), J. Eric Ivancich)
- rgw: rgw/iam: correcting the result of get role policy ([pr#36193](#), Pritha Srivastava)

- rgw: rgw/url: fix amqp urls with vhosts ([pr#35384](#), Yuval Lifshitz)
- rgw: stop realm reloader before store shutdown ([pr#36192](#), Casey Bodley)
- tools: Add statfs operation to ceph-objectstore-tool ([pr#35713](#), David Zafman)

## v14.2.10 Nautilus

This is the tenth release in the Nautilus series. In addition to fixing a security-related bug in RGW, this release brings a number of bugfixes across all major components of Ceph. We recommend that all Nautilus users upgrade to this release.

## Notable Changes

- CVE-2020-10753: rgw: sanitize newlines in s3 CORSConfiguration's ExposeHeader ([William Bowling](#), [Adam Mohammed](#), [Casey Bodley](#))
- RGW: Bucket notifications now support Kafka endpoints. This requires librkdafka of version 0.9.2 and up. Note that Ubuntu 16.04.6 LTS (Xenial Xerus) has an older version of librkdafka, and would require an update to the library.
- The pool parameter `target_size_ratio`, used by the pg autoscaler, has changed meaning. It is now normalized across pools, rather than specifying an absolute ratio. For details, see [Autoscaling placement groups](#). If you have set target size ratios on any pools, you may want to set these pools to autoscale `warn` mode to avoid data movement during the upgrade:

```
1. ceph osd pool set <pool-name> pg_autoscale_mode warn
```

- The behaviour of the `-o` argument to the rados tool has been reverted to its original behaviour of indicating an output file. This reverts it to a more consistent behaviour when compared to other tools. Specifying object size is now accomplished by using an upper case O `-O`.
- The format of MDSs in ceph fs dump has changed.
- Ceph will issue a health warning if a RADOS pool's `size` is set to 1 or in other words the pool is configured with no redundancy. This can be fixed by setting the pool size to the minimum recommended value with:

```
1. ceph osd pool set <pool-name> size <num-replicas>
```

The warning can be silenced with:

```
1. ceph config set global mon_warn_on_pool_no_redundancy false
```

- RGW: bucket listing performance on sharded bucket indexes has been notably improved by heuristically – and significantly, in many cases – reducing the number of entries requested from each bucket index shard.

## Changelog

---

- build/ops: address SELinux denials observed in rgw/multisite test run ([pr#34539](#), Kefu Chai, Kaleb S. Keithley)
- build/ops: ceph.spec.in: build on el8 ([pr#35599](#), Kefu Chai, Brad Hubbard, Alfonso Martínez, Nathan Cutler, Sage Weil, luo.runbing)
- build/ops: cmake: Improve test for 16-byte atomic support on IBM Z ([pr#33716](#), Ulrich Weigand)
- build/ops: do\_cmake.sh: fix application of -DWITH\_RADOSGW\_KAFKA\_ENDPOINT=OFF ([pr#34008](#), Nathan Cutler, Kefu Chai)
- build/ops: install-deps.sh: Use dnf for rhel/centos 8 ([pr#35461](#), Brad Hubbard)
- build/ops: rpm: add python3-saml as install dependency ([pr#34475](#), Kefu Chai, Ernesto Puerta)
- build/ops: selinux: Allow ceph to setsched ([pr#34433](#), Brad Hubbard)
- build/ops: selinux: Allow ceph-mgr access to httpd dir ([pr#34434](#), Brad Hubbard)
- build/ops: selinux: Allow getattr access to /proc/kcore ([pr#34870](#), Brad Hubbard)
- build/ops: spec: address some warnings raised by RPM 4.15.1 ([pr#34527](#), Nathan Cutler)
- ceph-volume/batch: check lvs list before access ([pr#34481](#), Jan Fajerski)
- ceph-volume/batch: return success when all devices are filtered ([pr#34478](#), Jan Fajerski)
- ceph-volume: add and delete lvm tags in a single lvchange call ([pr#35453](#), Jan Fajerski)
- ceph-volume: add ceph.osdspec\_affinity tag ([pr#35132](#), Joshua Schmid)
- ceph-volume: devices/simple/scan: Fix string in log statement ([pr#34445](#), Jan Fajerski)
- ceph-volume: fix nautilus functional tests ([pr#33391](#), Jan Fajerski)
- ceph-volume: lvm: get\_device\_vgs() filter by provided prefix ([pr#33616](#), Jan Fajerski, Yehuda Sadeh)
- ceph-volume: prepare: use \*-slots arguments for implicit sizing ([pr#34278](#), Jan

Fajerski)

- ceph-volume: silence ‘ceph-bluestore-tool’ failures ([pr#33428](#), Sébastien Han)
- ceph-volume: strip \_dmcrypt suffix in simple scan json output ([pr#33722](#), Jan Fajerski)
- cephfs/tools: add accounted\_rstat/rstat when building file dentry ([pr#35185](#), Xiubo Li)
- cephfs/tools: cephfs-journal-tool: correctly parse –dry\_run argument ([pr#34784](#), Milind Changire)
- cephfs: allow pool names with hyphen and period ([pr#35391](#), Rishabh Dave, Ramana Raja)
- cephfs: ceph-fuse: link to libfuse3 and pass “-o big\_writes” to libfuse if libfuse < 3.0.0 ([pr#34771](#), Kefu Chai, Xiubo Li, “Yan, Zheng”)
- cephfs: client: expose Client::ll\_register\_callback via libcephfs ([pr#35393](#), Kefu Chai, Jeff Layton)
- cephfs: client: fix Finisher assert failure ([pr#35000](#), Xiubo Li)
- cephfs: client: fix bad error handling in lseek SEEK\_HOLE / SEEK\_DATA ([pr#34308](#), Jeff Layton)
- cephfs: client: only set MClientCaps::FLAG\_SYNC when flushing dirty auth caps ([pr#35118](#), Jeff Layton)
- cephfs: client: reset requested\_max\_size if file write is not wanted ([pr#34767](#), “Yan, Zheng”)
- cephfs: mds: Handle blacklisted error in purge queue ([pr#35149](#), Varsha Rao)
- cephfs: mds: SIGSEGV in Migrator::export\_sessions\_flushed ([pr#33751](#), “Yan, Zheng”)
- cephfs: mds: Using begin() and empty() to iterate the xlist ([pr#34338](#), Shen Hang, “Yan, Zheng”)
- cephfs: mds: add configurable snapshot limit ([pr#33295](#), Milind Changire)
- cephfs: mds: display scrub status in ceph status ([issue#41508](#), [issue#42713](#), [issue#44520](#), [issue#42168](#), [issue#42169](#), [issue#42569](#), [issue#41424](#), [issue#42835](#), [issue#36370](#), [issue#42325](#), [pr#30704](#), Venky Shankar, Patrick Donnelly, Sage Weil, Kefu Chai)
- cephfs: mds: don’t shallow copy when decoding xattr map ([pr#35199](#), “Yan, Zheng”)
- cephfs: mds: handle bad purge queue item encoding ([pr#34307](#), “Yan, Zheng”)

- cephfs: mds: handle ceph\_assert on blacklisting ([pr#34435](#), Milind Changire)
- cephfs: mds: just delete MDSIOContextBase during shutdown ([pr#34343](#), "Yan, Zheng", Patrick Donnelly)
- cephfs: mds: take xlock in the order requests start locking ([pr#35392](#), "Yan, Zheng")
- common/bl: fix memory corruption in bufferlist::claim\_append() ([pr#34516](#), Radoslaw Zarzynski)
- common/blkdev: compilation of telemetry and device backports ([pr#33726](#), Sage Weil, Difan Zhang, Patrick Seidensal, Kefu Chai)
- common/blkdev: fix some problems with smart scraping ([pr#33421](#), Sage Weil)
- common/ceph\_time: tolerate mono time going backwards ([pr#34542](#), Sage Weil)
- common/options: Disable bluefs\_buffered\_io by default again ([pr#34297](#), Mark Nelson)
- compressor/lz4: work around bug in liblz4 versions <1.8.2 ([pr#35004](#), Sage Weil, Dan van der Ster)
- core: bluestore/bdev: initialize size when creating object ([pr#34832](#), Willem Jan Withagen)
- core: bluestore: Don't pollute old journal when add new device ([pr#34796](#), Yang Honggang)
- core: bluestore: fix 'unused' calculation ([pr#34794](#), xie xingguo, Igor Fedotov)
- core: bluestore: fix extent leak after main device expand ([pr#34711](#), Igor Fedotov)
- core: bluestore: more flexible DB volume space usage ([pr#33889](#), Igor Fedotov)
- core: bluestore: open DB in read-only when expanding DB/WAL ([pr#34611](#), Igor Fedotov, Jianpeng Ma, Adam Kupczyk)
- core: bluestore: prevent BlueFS::dirty\_files from being leaked when syncing metadata ([pr#34515](#), Xuehan Xu)
- core: msg/async/rdma: fix bug event center is blocked by rdma construct connection for transport ib sync msg ([pr#34780](#), Peng Liu)
- core: msgr: backport the EventCenter-related fixes ([pr#33820](#), Radoslaw Zarzynski, Jeff Layton, Kefu Chai)
- core: rados: prevent ShardedOpWQ suicide\_grace drop when waiting for work ([pr#34882](#), Dan Hill)

- doc/mgr/telemetry: added device channel details ([pr#33684](#), Yaarit Hatuka)
- doc/releases/nautilus: restart OSDs to make them bind to v2 addr ([pr#34524](#), Nathan Cutler)
- doc: fix parameter to set pg autoscale mode ([pr#34518](#), Changcheng Liu)
- doc: mds-config-ref: update 'mds\_log\_max\_segments' value ([pr#35278](#), Konstantin Shalygin)
- doc: reset PendingReleaseNotes following 14.2.8 release ([pr#33863](#), Nathan Cutler)
- global: ensure CEPH\_ARGS is decoded before early arg processing ([pr#33261](#), Kefu Chai, Jason Dillaman)
- mgr/DaemonServer: fix pg merge checks ([pr#34354](#), Sage Weil)
- mgr/PyModule: fix missing tracebacks in handle\_pyerror() ([pr#34627](#), Tim Serong)
- mgr/balancer: tolerate pgs outside of target weight map ([pr#34761](#), Sage Weil)
- mgr/dashboard/grafana: Add rbd-image details dashboard ([pr#35248](#), Enno Gotthold)
- mgr/dashboard: 'destroyed' view in CRUSH map viewer ([pr#33764](#), Avan Thakkar)
- mgr/dashboard: Add more debug information to Dashboard RGW backend ([pr#34399](#), Volker Theile)
- mgr/dashboard: Dashboard does not allow you to set norebalance OSD flag ([pr#33927](#), Nizamudeen)
- mgr/dashboard: Disable cache for static files ([pr#33763](#), Tiago Melo)
- mgr/dashboard: Display the aggregated number of request ([pr#35212](#), Tiago Melo)
- mgr/dashboard: Fix HomeTest setup ([pr#35086](#), Tiago Melo)
- mgr/dashboard: Fix cherrypy request logging error ([pr#31586](#), Kiefer Chang)
- mgr/dashboard: Fix error in unit test caused by timezone ([pr#34473](#), Tiago Melo)
- mgr/dashboard: Fix error when listing RBD while deleting or moving ([pr#34120](#), Tiago Melo)
- mgr/dashboard: Fix iSCSI's username and password validation ([pr#34550](#), Tiago Melo)
- mgr/dashboard: Fixes rbd image 'purge trash' button & modal text ([pr#33697](#), anurag)
- mgr/dashboard: Improve workaround to redraw datatables ([pr#34413](#), Volker Theile)
- mgr/dashboard: Not able to restrict bucket creation for new user ([pr#34692](#),

Volker Theile)

- mgr/dashboard: Pool read/write OPS shows too many decimal places ([pr#34039](#), anurag, Ernesto Puerta)
- mgr/dashboard: Prevent iSCSI target recreation when editing controls ([pr#34551](#), Tiago Melo)
- mgr/dashboard: REST API: OpenAPI docs require internet connection ([pr#33032](#), Patrick Seidensal)
- mgr/dashboard: RGW port autodetection does not support "Beast" RGW frontend ([pr#34400](#), Volker Theile)
- mgr/dashboard: Refactor Python unittests and controller ([pr#34662](#), Volker Theile)
- mgr/dashboard: Repair broken grafana panels ([pr#34417](#), Kristoffer Grönlund)
- mgr/dashboard: Searchable objects for table ([pr#32891](#), Stephan Müller)
- mgr/dashboard: Tabs does not handle click events ([issue#39326](#), [pr#34282](#), Tiago Melo)
- mgr/dashboard: UI fixes ([pr#34038](#), Avan Thakkar)
- mgr/dashboard: Updated existing E2E tests to match new format ([pr#33024](#), Nathan Weinberg)
- mgr/dashboard: Use booleanText pipe ([pr#33234](#), Alfonso Martínez, Volker Theile)
- mgr/dashboard: Use default language when running "npm run build" ([pr#33668](#), Tiago Melo)
- mgr/dashboard: do not show RGW API keys if only read-only privileges ([pr#33665](#), Alfonso Martínez)
- mgr/dashboard: fix COVERAGE\_PATH in run-backend-api-tests.sh ([pr#34489](#), Alfonso Martínez)
- mgr/dashboard: fix backport #33764 ([pr#34640](#), Ernesto Puerta)
- mgr/dashboard: fix error when enabling SSO with cert. file ([pr#34129](#), Alfonso Martínez)
- mgr/dashboard: fix py2 strftime ImportError (not thread safe) ([pr#35016](#), Alfonso Martínez)
- mgr/dashboard: fixing RBD purge error in backend ([pr#34847](#), Kiefer Chang)
- mgr/dashboard: install teuthology using pip ([pr#35174](#), Nathan Cutler, Kefu Chai)
- mgr/dashboard: list configured prometheus alerts ([pr#34373](#), Patrick Seidensal,

Tiago Melo)

- mgr/dashboard: monitoring menu entry should indicate firing alerts ([pr#34823](#), Tiago Melo, Volker Theile)
- mgr/dashboard: remove 'config-opt: read' perm. from system roles ([pr#33739](#), Alfonso Martínez)
- mgr/dashboard: show checkboxes for booleans ([pr#33388](#), Tatjana Dehler)
- mgr/dashboard: use FQDN for failover redirection ([pr#34497](#), Ernesto Puerta)
- mgr/insights: fix prune-health-history ([pr#35214](#), Sage Weil)
- mgr/pg\_autoscaler: fix division by zero ([pr#33420](#), Sage Weil)
- mgr/pg\_autoscaler: treat target ratios as weights ([pr#34087](#), Josh Durgin)
- mgr/prometheus: ceph\_pg\_\* metrics contains last value instead of sum across all reported states ([pr#34162](#), Jacek Suchenia)
- mgr/run-tox-tests: Fix issue with PYTHONPATH ([pr#33688](#), Brad Hubbard)
- mgr/telegraf: catch FileNotFoundError exception ([pr#34628](#), Kefu Chai)
- mgr/telemetry: add 'last\_upload' to status ([pr#33409](#), Yaarit Hatuka)
- mgr/telemetry: catch exception during requests.put ([pr#33141](#), Sage Weil)
- mgr/telemetry: fix UUID and STR concat ([pr#33666](#), Yaarit Hatuka)
- mgr/telemetry: fix and document proxy usage ([pr#33649](#), Lars Marowsky-Bree)
- mgr/volumes: Add interface to get subvolume metadata ([pr#34679](#), Kotresh HR)
- mgr/volumes: fs subvolume clone cancel ([issue#44208](#), [pr#34036](#), Venky Shankar, Michael Fritch)
- mgr/volumes: minor fixes ([pr#35482](#), Kotresh HR)
- mgr/volumes: synchronize ownership (for symlinks) and inode timestamps for cloned subvolumes ([issue#24880](#), [issue#43965](#), [pr#33877](#), Ramana Raja, Rishabh Dave, huanwen ren, Venky Shankar, Jos Collin)
- mgr: Add get\_rates\_from\_data to mgr\_util.py ([pr#33893](#), Stephan Müller, Ernesto Puerta)
- mgr: Improve internal python to c++ interface ([pr#34356](#), David Zafman)
- mgr: close restful socket after exec ([pr#35213](#), liushi)
- mgr: force purge normal ceph entities from service map ([issue#44677](#), [pr#34563](#), Venky Shankar)

- mgr: synchronize ClusterState's health and mon\_status ([pr#34326](#), Radoslaw Zarzynski)
- mgr: update "hostname" when we already have the daemon state from that entity ([pr#33834](#), Kefu Chai)
- mon/FSCommands: Fix 'add\_data\_pool' command and 'fs new' command ([pr#34774](#), Ramana Raja)
- mon/OSDMonitor: Always tune priority cache manager memory on all mons ([pr#34916](#), Sridhar Seshasayee)
- mon/OSDMonitor: allow trimming maps even if osds are down ([pr#34983](#), Joao Eduardo Luis)
- mon/PGMap: fix summary display of >32bit pg states ([pr#33275](#), Sage Weil, Adam C. Emerson)
- mon: Get session\_map\_lock before remove\_session ([pr#34677](#), Xiaofei Cui)
- mon: calculate min\_size on osd pool set size ([pr#34585](#), Deepika Upadhyay)
- mon: disable min pg per osd warning ([pr#34618](#), Sage Weil)
- mon: fix/improve mon sync over small keys ([pr#33765](#), Sage Weil)
- mon: stash newer map on bootstrap when addr doesn't match ([pr#34500](#), Sage Weil)
- monitoring: Fix "10% OSDs down" alert description ([pr#35211](#), Benoît Knecht)
- monitoring: Fix pool capacity incorrect ([pr#34450](#), James Cheng)
- monitoring: alert for pool fill up broken ([pr#35137](#), Volker Theile)
- monitoring: alert for prediction of disk and pool fill up broken ([pr#34394](#), Patrick Seidensal)
- monitoring: fix RGW grafana chart 'Average GET/PUT Latencies' ([pr#33860](#), Alfonso Martínez)
- monitoring: fix decimal precision in Grafana %percentages ([pr#34829](#), Ernesto Puerta)
- monitoring: root volume full alert fires false positives ([pr#34419](#), Patrick Seidensal)
- osd/OSD: Log slow ops/types to cluster logs ([pr#33503](#), Sage Weil, Sridhar Seshasayee)
- osd/OSDMap: Show health warning if a pool is configured with size 1 ([pr#31842](#), Sridhar Seshasayee)

- osd/PeeringState.h: ignore RemoteBackfillReserved in WaitLocalBackfillReserved ([pr#34512](#), Neha Ojha)
- osd/PeeringState: do not trim pg log past last\_update\_ondisk ([pr#34957](#), Samuel Just, xie xingguo)
- osd/PeeringState: transit async\_recovery\_targets back into acting before backfilling ([pr#32849](#), xie xingguo)
- osd: dispatch\_context and queue split finish on early bail-out ([pr#35024](#), Sage Weil)
- osd: fix racy accesses to OSD::osdmap ([pr#33530](#), Radoslaw Zarzynski)
- pybind/mgr/\*: fix config\_notify handling of default values ([pr#34116](#), Nathan Cutler, Sage Weil)
- pybind/mgr: use six==1.14.0 ([pr#34316](#), Kefu Chai)
- pybind/rbd: RBD.create() method's 'old\_format' parameter now defaults to False ([pr#35183](#), Jason Dillaman)
- pybind/rbd: ensure image is open before permitting operations ([pr#34424](#), Mykola Golub)
- pybind/rbd: fix no lockers are obtained, ImageNotFound exception will be output ([pr#34388](#), zhangdaolong)
- rbd: librbd: copy API should not inherit v1 image format by default ([pr#35182](#), Jason Dillaman)
- rbd: rbd-mirror: improve detection of blacklisted state ([pr#33533](#), Mykola Golub)
- rgw/kafka: add kafka endpoint support ([pr#32960](#), Yuval Lifshitz, Willem Jan Withagen, Kefu Chai)
- rgw/notifications: backporting features and bug fix ([pr#34107](#), Yuval Lifshitz)
- rgw/notifications: fix topic action fail with "MethodNotAllowed" ([issue#44614](#), [pr#33978](#), Yuval Lifshitz)
- rgw/notifications: version id was not sent in versioned buckets ([pr#35181](#), Yuval Lifshitz)
- rgw: when you abort a multipart upload request, the quota may be not updated ([pr#33268](#), Richard Bai(白学余))
- rgw: Add support bucket policy for subuser ([pr#33714](#), Seena Fallah)
- rgw: Fix dynamic resharding not working for empty zonegroup in period ([pr#33266](#), Or Friedmann)

- rgw: Fix upload part copy range able to get almost any string ([pr#33265](#), Or Friedmann)
- rgw: GET/HEAD and PUT operations on buckets w/lifecycle expiration configured do not return x-amz-expiration header ([pr#32924](#), Matt Benjamin, Yuval Lifshitz)
- rgw: MultipartObjectProcessor supports stripe size > chunk size ([pr#33271](#), Casey Bodley)
- rgw: ReplaceKeyPrefixWith and ReplaceKeyWith can not set at the same ... ([pr#34599](#), yuliyang)
- rgw: anonymous swift to obj that dont exist should 401 ([pr#35045](#), Matthew Oliver)
- rgw: clear ent\_list for each loop of bucket list ([issue#44394](#), [pr#34099](#), Yao Zongyou)
- rgw: dmclock: wait until the request is handled ([pr#34954](#), GaryHyg)
- rgw: find oldest period and update RGWMetadataLogHistory() ([pr#34597](#), Shilpa Jagannath)
- rgw: fix SignatureDoesNotMatch when use ipv6 address in s3 client ([pr#33267](#), yuliyang)
- rgw: fix bug with (un)ordered bucket listing and marker w/ namespace ([pr#34609](#), J. Eric Ivancich)
- rgw: fix lc does not delete objects that do not have exactly the same tags as the rule ([pr#35002](#), Or Friedmann)
- rgw: fix multipart upload's error response ([pr#35019](#), GaryHyg)
- rgw: fix rgw crash when duration is invalid in sts request ([pr#33273](#), yuliyang)
- rgw: fix some list buckets handle leak ([pr#34986](#), Tianshan Qu)
- rgw: get barbican secret key request maybe return error code ([pr#33965](#), Richard Bai(白学余))
- rgw: increase log level for same or older period pull msg ([pr#34833](#), Ali Maredia)
- rgw: make max\_connections configurable in beast ([pr#33340](#), Tiago Pasqualini)
- rgw: making implicit\_tenants backwards compatible ([issue#24348](#), [pr#33749](#), Marcus Watts)
- rgw: multisite: enforce spawn window for incremental data sync ([pr#33270](#), Casey Bodley)
- rgw: radosgw-admin: add support for -bucket-id in bucket stats command ([pr#34815](#), Vikhyat Umrao)

- rgw: radosgw-admin: fix infinite loops in 'datalog list' ([pr#35001](#), Casey Bodley)
- rgw: reshards: skip stale bucket id entries from reshards queue ([pr#34735](#), Abhishek Lekshmanan)
- rgw: set bucket attr twice when delete lifecycle config ([pr#34598](#), zhang Shaowen)
- rgw: set correct storage class for append ([pr#34064](#), yuliyang)
- rgw: sts: add all http args to req\_info ([pr#33355](#), yuliyang)
- rgw: tune sharded bucket listing ([pr#33675](#), J. Eric Ivancich)
- tests: migrate qa/ to python3 ([pr#34171](#), Kefu Chai, Sage Weil, Casey Bodley, Rishabh Dave, Patrick Donnelly, Kyr Shatskyy, Michael Fritch, Xiubo Li, Ilya Dryomov, Alfonso Martínez, Thomas Bechtold)
- tools/cli: bash\_completion: Do not auto complete obsolete and hidden cmds ([pr#35117](#), Kotresh HR)
- tools/cli: ceph\_argparse: increment matchcnt on kwargs ([pr#33160](#), Matthew Oliver, Shyukri Shyukriev)
- tools/rados: Unmask '-o' to restore original behaviour ([pr#33641](#), Brad Hubbard)

## v14.2.9 Nautilus

---

This is the ninth bugfix release of Nautilus. This release fixes a couple of security issues in RGW & Messenger V2. We recommend all users to upgrade to this release.

### Notable Changes

---

- CVE-2020-1759: Fixed nonce reuse in msgr V2 secure mode
- CVE-2020-1760: Fixed XSS due to RGW GetObject header-splitting

## v14.2.8 Nautilus

---

This is the eighth update to the Ceph Nautilus release series. This release fixes issues across a range of subsystems. We recommend that all users upgrade to this release.

### Notable Changes

---

- The default value of `bluestore_min_alloc_size_ssd` has been changed to 4K to improve performance across all workloads.

- The following OSD memory config options related to bluestore cache autotuning can now be configured during runtime:

```

    • osd_memory_base (default: 768 MB)
    • osd_memory_cache_min (default: 128 MB)
    • osd_memory_expected_fragmentation (default: 0.15)
    • osd_memory_target (default: 4 GB)

```

The above options can be set with:

- `ceph config set osd <option> <value>`

- The MGR now accepts `profile rbd` and `profile rbd-read-only` user caps. These caps can be used to provide users access to MGR-based RBD functionality such as `rbd perf image iostat` and `rbd perf image iotop`.
- The configuration value `osd_calc_pg_upmaps_max_stddev` used for upmap balancing has been removed. Instead use the mgr balancer config `upmap_max_deviation` which now is an integer number of PGs of deviation from the target PGs per OSD. This can be set with a command like `ceph config set mgr mgr/balancer/upmap_max_deviation 2`. The default `upmap_max_deviation` is 5. There are situations where crush rules would not allow a pool to ever have completely balanced PGs. For example, if crush requires 1 replica on each of 3 racks, but there are fewer OSDs in 1 of the racks. In those cases, the configuration value can be increased.
- RGW: a mismatch between the bucket notification documentation and the actual message format was fixed. This means that any endpoints receiving bucket notification, will now receive the same notifications inside a JSON array named 'Records'. Note that this does not affect pulling bucket notification from a subscription in a 'pubsub' zone, as these are already wrapped inside that array.
- CephFS: multiple active MDS forward scrub is now rejected. Scrub currently only is permitted on a file system with a single rank. Reduce the ranks to one via `ceph fs set <fs_name> max_mds 1`.
- Ceph now refuses to create a file system with a default EC data pool. For further explanation, see: <https://docs.ceph.com/docs/nautilus/cephfs/createfs/#creating-pools>
- Ceph will now issue a health warning if a RADOS pool has a `pg_num` value that is not a power of two. This can be fixed by adjusting the pool to a nearby power of two:

- `ceph osd pool set <pool-name> pg_num <new-pg-num>`

Alternatively, the warning can be silenced with:

```
1. ceph config set global mon_warn_on_pool_pg_num_not_power_of_two false
```

## Changelog

- bluestore: common/options: bluestore 4k min\_alloc\_size for SSD ([pr#32998](#), Mark Nelson, Sage Weil)
- bluestore: os/bluestore: Add config observer for osd memory specific options ([pr#31852](#), Sridhar Seshasayee)
- bluestore: os/bluestore/BlueStore.cc: set priorities for compression stats ([pr#32845](#), Neha Ojha)
- bluestore: os/bluestore: default bluestore\_block\_size 1T -> 100G ([pr#32283](#), Sage Weil)
- build/ops: cmake: remove seastar tests from “make check” ([pr#32658](#), Kefu Chai)
- build/ops: install-deps,rpm: enable devtoolset-8 on aarch64 also ([issue#38892](#), [pr#32651](#), Kefu Chai)
- build/ops: rpm: add rpm-build to SUSE-specific make check deps ([pr#32208](#), Nathan Cutler)
- build/ops: switch to boost 1.72 ([pr#32441](#), Willem Jan Withagen, Kefu Chai)
- build/ops: tools/setup-virtualenv.sh: do not default to python2.7 ([pr#30739](#), Nathan Cutler)
- cephfs: cephfs-journal-tool: fix crash and usage ([pr#32913](#), Xiubo Li)
- cephfs: client: Add is\_dir() check before changing directory ([pr#32916](#), Varsha Rao)
- cephfs: client: add procession of SEEK\_HOLE and SEEK\_DATA in lseek ([pr#30764](#), Shen Hang)
- cephfs: client: add warning when cap != in->auth\_cap ([pr#32065](#), Shen Hang)
- cephfs: client: EINVAL may be returned when offset is 0 ([pr#30762](#), wenpengLi)
- cephfs: client: fix lazyio\_synchronize() to update file size and libcephfs: Add Tests for LazyIO ([pr#30769](#), Sidharth Anupkrishnan)
- cephfs: client: \_readdir\_cache\_cb() may use the readdir\_cache already clear ([issue#41148](#), [pr#30763](#), huanwen ren)
- cephfs: client: remove Inode.dir\_contacts field and handle bad whence value to

- llseek gracefully ([pr#30766](#), Jeff Layton)
- cephfs,common: osdc/objecter: Fix last\_sent in scientific format and add age to ops ([pr#31081](#), Varsha Rao)
- cephfs: disallow changing fuse\_default\_permissions option at runtime ([pr#32915](#), Zhi Zhang)
- cephfs: mds: add command that config individual client session ([issue#40811](#), [pr#32245](#), "Yan, Zheng")
- cephfs: mds: "apply configuration changes through MDSRank" and "recall caps from quiescent sessions" and "drive cap recall while dropping cache" ([pr#30761](#), Patrick Donnelly, Jeff Layton)
- cephfs: mds: fix assert(omap\_num\_objs <= MAX\_OBJECTS) of OpenFileTable ([pr#32756](#), "Yan, Zheng")
- cephfs: mds: fix revoking caps after stale->resume circle ([pr#32909](#), "Yan, Zheng")
- cephfs: mds: free heap memory may grow too large for some workloads ([pr#31802](#), Patrick Donnelly)
- cephfs: MDSMonitor: warn if a new file system is being created with an EC default data pool ([pr#32600](#), Patrick Donnelly)
- cephfs: mds: no assert on frozen dir when scrub path ([pr#32071](#), Zhi Zhang)
- cephfs: mds: note client features when rejecting client ([pr#32914](#), Patrick Donnelly)
- cephfs: mds/OpenFileTable: match MAX\_ITEMS\_PER\_OBJ to osd\_deep\_scrub\_large\_omap\_object\_key\_threshold ([pr#32921](#), Vikhyat Umrao, Varsha Rao)
- cephfs: mds: properly evaluate unstable locks when evicting client ([pr#32073](#), "Yan, Zheng")
- cephfs: mds: reject forward scrubs when cluster has multiple active MDS (more than one rank) ([pr#32602](#), Patrick Donnelly, Milind Changire)
- cephfs: mds: reject sessionless messages ([issue#40784](#), [pr#30843](#), "Yan, Zheng", Xiao Guodong, Shen Hang)
- cephfs: mds: remove unnecessary debug warning ([pr#32077](#), Patrick Donnelly)
- cephfs: mds returns -5(EIO) error when the deleted file does not exist ([pr#30767](#), huanwen ren)
- cephfs: mds: split the dir if the op makes it oversized, because some ops maybe

- in flight ([pr#31302](#), simon gao)
- cephfs: mds: tolerate no snaprealm encoded in on-disk root inode ([pr#32079](#), "Yan, Zheng")
  - cephfs: mgr: "mds metadata" to setup new DaemonState races with fsmap ([pr#31905](#), Patrick Donnelly)
  - cephfs: mgr/volumes: allow setting uid, gid of subvolume and subvolume group during creation ([issue#42923](#), [pr#31741](#), Venky Shankar, Jos Collin)
  - cephfs: mgr/volumes: fetch trash and clone entries without blocking volume access ([issue#44282](#), [pr#33526](#), Venky Shankar)
  - cephfs: mgr/volumes: fs subvolume resize command ([pr#31332](#), Jos Collin)
  - cephfs: mgr/volumes: misc fix and feature enhancements ([issue#42646](#), [issue#43645](#), [pr#33122](#), Rishabh Dave, Joshua Schmid, Venky Shankar, Ramana Raja, Jos Collin)
  - cephfs: mgr/volumes: unregister job upon async threads exception ([issue#44315](#), [pr#33569](#), Venky Shankar)
  - cephfs: mon: print FSMap regardless of file system count ([pr#32912](#), Patrick Donnelly)
  - cephfs: pybind/mgr/volumes: idle connection drop is not working ([pr#33116](#), Patrick Donnelly)
  - cephfs: RuntimeError: Files in flight high water is unexpectedly low (0 / 6) ([pr#33115](#), Patrick Donnelly)
  - ceph.in: check ceph-conf returncode ([pr#31367](#), Dimitri Savineau)
  - ceph-monstore-tool: correct the key for storing mgr\_command\_descs ([pr#33278](#), Kefu Chai)
  - ceph-volume: add db and wal support to raw mode ([pr#32979](#), Sébastien Han)
  - ceph-volume: add methods to pass filters to pvs, vgs and lvs commands ([pr#33217](#), Rishabh Dave)
  - ceph-volume: add raw (-bluestore) mode ([pr#32733](#), Jan Fajerski, Sage Weil)
  - ceph-volume: add sizing arguments to prepare ([pr#33231](#), Jan Fajerski)
  - ceph-volume: allow raw block devices everywhere ([pr#32868](#), Jan Fajerski)
  - ceph-volume: assume msgrV1 for all branches containing mimic ([pr#31616](#), Jan Fajerski)
  - ceph-volume: avoid calling zap\_lv with a LV-less VG ([pr#33297](#), Jan Fajerski)

- ceph-volume: batch bluestore fix create\_lvs call ([pr#33232](#), Jan Fajerski)
- ceph-volume: batch bluestore fix create\_lvs call ([pr#33301](#), Jan Fajerski)
- ceph-volume/batch: fail on filtered devices when non-interactive ([pr#33202](#), Jan Fajerski)
- ceph-volume: Dereference symlink in lvm list ([pr#32877](#), Benoît Knecht)
- ceph-volume: don't remove vg twice when zapping filestore ([pr#33337](#), Jan Fajerski)
- ceph-volume: finer grained availability notion in inventory ([pr#33240](#), Jan Fajerski)
- ceph-volume: fix has\_bluestore\_label() function ([pr#33239](#), Guillaume Abrioux)
- ceph-volume: fix is\_ceph\_device for lvm batch ([pr#33253](#), Jan Fajerski, Dimitri Savineau)
- ceph-volume: fix the integer overflow ([pr#32873](#), dongdong tao)
- ceph-volume: import mock.mock instead of unittest.mock (py2) ([pr#32870](#), Jan Fajerski)
- ceph-volume/lvm/activate.py: clarify error message: fsid refers to osd\_fsid ([pr#32864](#), Yaniv Kaul)
- ceph-volume: lvm/deactivate: add unit tests, remove -all ([pr#32863](#), Jan Fajerski)
- ceph-volume: lvm deactivate command ([pr#33209](#), Jan Fajerski)
- ceph-volume: make get\_devices fs location independent ([pr#33200](#), Jan Fajerski)
- ceph-volume: minor clean-up of "simple scan" subcommand help ([pr#32556](#), Michael Fritch)
- ceph-volume: pass journal\_size as Size not string ([pr#33334](#), Jan Fajerski)
- ceph-volume: refactor listing.py + fixes ([pr#33238](#), Jan Fajerski, Rishabh Dave, Guillaume Abrioux)
- ceph-volume: reject disks smaller than 5GB in inventory ([issue#40776](#), [pr#31554](#), Jan Fajerski)
- ceph-volume: skip osd creation when already done ([pr#33242](#), Guillaume Abrioux)
- ceph-volume/test: patch VolumeGroups ([pr#32558](#), Jan Fajerski)
- ceph-volume: use correct extents if using db-devices and >1 osds\_per\_device ([pr#32874](#), Fabian Niepelt)

- ceph-volume: use fsync for dd command ([pr#31553](#), Rishabh Dave)
- ceph-volume: use get\_device\_vgs in has\_common\_vg ([pr#33254](#), Jan Fajerski)
- ceph-volume: util: look for executable in \$PATH ([pr#32860](#), Shyukri Shyukriev)
- ceph-volume/zfs: add the inventory command ([pr#31295](#), Willem Jan Withagen)
- common/admin\_socket: Increase socket timeouts ([pr#32063](#), Brad Hubbard)
- common/bl: fix the dangling last\_p issue ([pr#33277](#), Radoslaw Zarzynski)
- common/config: update values when they are removed via mon ([pr#32846](#), Sage Weil)
- common: FIPS: audit and switch some memset & bzero users ([pr#32167](#), Radoslaw Zarzynski)
- common: fix deadlocky inflight op visiting in OpTracker ([pr#32858](#), Radoslaw Zarzynski)
- common/options: remove unused ms\_msgr2\_{sign, encrypt} ([pr#31850](#), Ilya Dryomov)
- common/util: use ifstream to read from /proc files ([pr#32901](#), Kefu Chai, songweibin)
- core: auth/Crypto: fallback to /dev/urandom if getentropy() fails ([pr#31301](#), Kefu Chai)
- core: mon: keep v1 address type when explicitly set ([pr#32028](#), Ricardo Dias)
- core: mon/OSDMonitor: Fix pool set target\_size\_bytes (etc) with unit suffix ([pr#31740](#), Prashant D)
- core: osd/OSDMap: health alert for non-power-of-two pg\_num ([pr#30689](#), Sage Weil)
- crush/CrushWrapper: behave with empty weight vector ([pr#32905](#), Kefu Chai)
- doc/cephfs/client-auth: description and example are inconsistent ([pr#32781](#), Ilya Dryomov)
- doc/cephfs: improve add/remove MDS section ([issue#39620](#), [pr#31116](#), Patrick Donnelly)
- doc/ceph-fuse: mention -k option in ceph-fuse man page ([pr#30765](#), Rishabh Dave)
- doc/ceph-volume: initial docs for zfs/inventory and zfs/api ([pr#32746](#), Willem Jan Withagen)
- doc: remove invalid option mon\_pg\_warn\_max\_per\_osd ([pr#31300](#), zhang daolong)
- doc/\_templates/page.html: redirect to etherpad ([pr#32248](#), Neha Ojha)
- doc: wrong datatype describing crush\_rule ([pr#32254](#), Kefu Chai)

- global: disable THP for Ceph daemons ([pr#31646](#), Patrick Donnelly, Mark Nelson)
- kv: fix shutdown vs async compaction ([pr#32715](#), Sage Weil)
- librbd: diff iterate with fast-diff now correctly includes parent ([pr#32469](#), Jason Dillaman)
- librbd: fix rbd\_open\_by\_id, rbd\_open\_by\_id\_read\_only ([pr#32837](#), yangjun)
- librbd: remove pool objects when removing a namespace ([pr#32839](#), Jason Dillaman)
- librbd: skip stale child with non-existent pool for list descendants ([pr#32841](#), songweibin)
- librbd: support compression allocation hints to the OSD ([pr#32842](#), Jason Dillaman)
- mgr: add 'rbd' profiles to support 'rbd\_support' module commands ([pr#32086](#), Jason Dillaman)
- mgr/alerts: simple health alerts ([pr#30820](#), Sage Weil)
- mgr: Balancer fixes ([pr#31956](#), Neha Ojha, Kefu Chai, David Zafman)
- mgr/DaemonServer: fix 'osd ok-to-stop' for EC pools ([pr#32844](#), Sage Weil)
- mgr/dashboard: add debug mode, and accept expected exception when SSL handshaking ([pr#31190](#), Kefu Chai, Ernesto Puerta, Joshua Schmid)
- mgr/dashboard: block mirroring page results in internal server error ([pr#32133](#), Jason Dillaman)
- mgr/dashboard: check embedded Grafana dashboard references ([issue#40008](#), [pr#31808](#), Kiefer Chang)
- mgr/dashboard: check if user has config-opt permissions ([pr#32827](#), Alfonso Martínez)
- mgr/dashboard: Cross sign button not working for some modals ([pr#32012](#), Ricardo Marques)
- mgr/dashboard: Dashboard can't handle self-signed cert on Grafana API ([pr#31792](#), Volker Theile)
- mgr/dashboard: disable 'Add Capability' button in rgw user edit ([pr#32930](#), Alfonso Martínez)
- mgr/dashboard: fix restored RBD image naming issue ([pr#31810](#), Kiefer Chang)
- mgr/dashboard: grafana charts match time picker selection ([pr#31999](#), Alfonso Martínez)

- mgr/dashboard, grafana: remove shortcut menu ([pr#31980](#), Ernesto Puerta)
- mgr/dashboard: Handle always-on Ceph Manager modules correctly ([pr#31782](#), Volker Theile)
- mgr/dashboard: Hardening accessing the metadata ([pr#32128](#), Volker Theile)
- mgr/dashboard: iSCSI targets not available if any gateway is down (and more...) ([pr#32304](#), Ricardo Marques)
- mgr/dashboard: KeyError on dashboard reload ([pr#32233](#), Patrick Seidensal)
- mgr/dashboard: key-value-table doesn't render booleans ([pr#31789](#), Patrick Seidensal)
- mgr/dashboard: Remove compression mode unset in pool from ([pr#31784](#), Stephan Müller)
- mgr/dashboard: show "Rename" in header & button when renaming RBD ([pr#31779](#), Alfonso Martínez)
- mgr/dashboard: sort monitors by open sessions correctly ([pr#31791](#), Alfonso Martínez)
- mgr/dashboard: Standby Dashboards don't handle all requests properly ([pr#32299](#), Volker Theile)
- mgr/dashboard: Trim IQN on iSCSI target form ([pr#31942](#), Ricardo Marques)
- mgr/dashboard: Unable to set boolean values to false when default is true ([pr#31941](#), Ricardo Marques)
- mgr/dashboard: Using wrong identifiers in RGW user/bucket datatables ([pr#32888](#), Volker Theile)
- mgr/devicehealth: ensure we don't store empty objects ([pr#31735](#), Sage Weil)
- mgr/devicehealth: fix telemetry stops sending device reports after 48 hours ([pr#33346](#), Yaarit Hatuka, Sage Weil)
- mgr: drop reference to msg on return ([pr#33498](#), Patrick Donnelly)
- mgr/MgrClient: fix open condition ([pr#32769](#), Sage Weil)
- mgr/pg\_autoscaler: calculate pool\_pg\_target using pool size ([pr#33170](#), Dan van der Ster)
- mgr/pg\_autoscaler: default to pg\_num[\_min] = 16 ([pr#32069](#), Sage Weil)
- mgr/pg\_autoscaler: default to pg\_num[\_min] = 32 ([pr#32931](#), Neha Ojha)
- mgr/pg\_autoscaler: implement shutdown method ([pr#32068](#), Patrick Donnelly)

- mgr/pg\_autoscaler: only generate target\_\* health warnings if targets set ([pr#32067](#), Sage Weil)
- mgr/prometheus: assign a value to osd\_dev\_node when obj\_store is not filestore or bluestore ([pr#31556](#), jiahuizeng)
- mgr/prometheus: report per-pool pg states ([pr#33157](#), Aleksei Zakharov)
- mgr/telemetry: anonymizing smartctl report itself ([pr#33082](#), Yaarit Hatuka)
- mgr/telemetry: check get\_metadata return val ([pr#33095](#), Yaarit Hatuka)
- mgr/telemetry: split entity\_name only once (handle ids with dots) ([pr#33168](#), Dan Mick)
- mgr/zabbix: Adds possibility to send data to multiple zabbix servers ([pr#30009](#), slivik, Jakub Sliva)
- mon/ConfigMonitor: fix handling of NO\_MON\_UPDATE settings ([pr#32856](#), Sage Weil)
- mon/ConfigMonitor: only propose if leader ([pr#33155](#), Sage Weil)
- mon: Don't put session during feature change ([pr#33152](#), Brad Hubbard)
- mon: elector: return after triggering a new election ([pr#33007](#), Greg Farnum)
- monitoring: wait before firing osd full alert ([pr#32070](#), Patrick Seidensal)
- mon/MgrMonitor.cc: add always\_on\_modules to the output of "ceph mgr module ls" ([pr#32997](#), Neha Ojha)
- mon/MgrMonitor.cc: warn about missing mgr in a cluster with osds ([pr#33142](#), Neha Ojha)
- mon/OSDMonitor: Don't update mon cache settings if rocksdb is not used ([pr#32520](#), Sridhar Seshasayee, Sage Weil)
- mon/OSDMonitor: fix format error ceph osd stat -format json ([pr#32062](#), Zheng Yin)
- mon/PGMap.h: disable network stats in dump\_osd\_stats ([pr#32466](#), Neha Ojha, David Zafman)
- mon: remove the restriction of address type in init\_with\_hosts ([pr#31844](#), Hao Xiong)
- mon/Session: only index osd ids >= 0 ([pr#32908](#), Sage Weil)
- mount.ceph: give a hint message when no mds is up or cluster is laggy ([pr#32910](#), Xiubo Li)
- mount.ceph: remove arbitrary limit on size of name= option ([pr#32807](#), Jeff Layton)

- msg: async/net\_handler.cc: Fix compilation ([pr#31736](#), Carlos Valiente)
- osd: add osd\_fast\_shutdown option (default true) ([pr#32743](#), Sage Weil)
- osd: Allow 64-char hostname to be added as the “host” in CRUSH ([pr#33147](#), Michal Skalski)
- osd: Diagnostic logging for upmap cleaning ([pr#32716](#), David Zafman)
- osd/OSD: enhance osd numa affinity compatibility ([pr#32843](#), luo rixin, Dai zhiwei)
- osd/PeeringState.cc: don’t let num\_objects become negative ([pr#32857](#), Neha Ojha)
- osd/PeeringState.cc: skip peer\_purged when discovering all missing ([pr#32847](#), Neha Ojha)
- osd/PeeringState: do not exclude up from acting\_recovery\_backfill ([pr#32064](#), Nathan Cutler, xie xingguo)
- osd/PrimaryLogPG: skip obcs that don’t exist during backfill scan\_range ([pr#31028](#), Sage Weil)
- osd: set affinity for \*all\* threads ([pr#31359](#), Sage Weil)
- osd: set collection pool opts on collection create, pg load ([pr#32123](#), Sage Weil)
- osd: Use physical ratio for nearfull (doesn’t include backfill resserve) ([pr#32773](#), David Zafman)
- pybind/mgr: Cancel output color control ([pr#31697](#), Zheng Yin)
- rbd: creating thick-provision image progress percent info exceeds 100% ([pr#32840](#), Xiangdong Mu)
- rbd: librbd: don’t call refresh from mirror::GetInfoRequest state machine ([pr#32900](#), Mykola Golub)
- rbd-mirror: clone v2 mirroring improvements ([pr#31518](#), Mykola Golub)
- rbd-mirror: fix ‘rbd mirror status’ asok command output ([pr#32447](#), Mykola Golub)
- rbd-mirror: make logrotate work ([pr#32593](#), Mykola Golub)
- rgw: add bucket permission verify when copy obj ([pr#31089](#), NancySu05)
- rgw: Adding ‘iam’ namespace for Role and User Policy related REST APIs ([pr#32437](#), Pritha Srivastava)
- rgw: adding mfa code validation when bucket versioning status is changed ([pr#32759](#), Pritha Srivastava)

- rgw: add num\_shards to radosgw-admin bucket stats ([pr#31182](#), Paul Emmerich)
- rgw: allow reshards log entries for non-existent buckets to be cancelled ([pr#32056](#), J. Eric Ivancich)
- rgw: auto-clean reshards queue entries for non-existent buckets ([pr#32055](#), J. Eric Ivancich)
- rgw: build\_linked\_oids\_for\_bucket and build\_buckets\_instance\_index should return negative value if it fails ([pr#32820](#), zhangshaowen)
- rgw: crypt: permit RGW-AUTO/default with SSE-S3 headers ([pr#31862](#), Matt Benjamin)
- rgw: data sync markers include timestamp from datalog entry ([pr#32819](#), Casey Bodley)
- rgw\_file: avoid string::front() on empty path ([pr#33008](#), Matt Benjamin)
- rgw: fix a bug that bucket instance obj can't be removed after resharding completed ([pr#32822](#), zhang Shaowen)
- rgw: fix an endless loop error when to show usage ([pr#31684](#), lvshuhua)
- rgw: fix bugs in listobjects v1 ([pr#32239](#), Albin Antony)
- rgw: fix compile errors with boost 1.70 ([pr#31289](#), Casey Bodley)
- rgw: fix data consistency error caused by rgw sent timeout ([pr#32821](#), 李纲彬82225)
- rgw: fix list versions starts with version\_id=null ([pr#30743](#), Tianshan Qu)
- rgw: fix one part of the bulk delete(RGWDeleteMultiObj\_ObjStore\_S3) fails but no error messages ([pr#33151](#), Snow Si)
- rgw: fix opslog operation field as per Amazon s3 ([issue#20978](#), [pr#32834](#), Jiaying Ren)
- rgw: fix refcount tags to match and update object's idtag ([pr#30741](#), J. Eric Ivancich)
- rgw: fix rgw crash when token is not base64 encode ([pr#32050](#), yuliyang)
- rgw: gc remove tag after all sub io finish ([issue#40903](#), [pr#30733](#), Tianshan Qu)
- rgw: Incorrectly calling ceph::buffer::list::decode\_base64 in bucket policy ([pr#32832](#), GaryHyg)
- rgw: maybe coredump when reload operator happened ([pr#33149](#), Richard Bai(白学余))
- rgw: move forward marker even in case of many rgw.none indexes ([pr#32824](#), Ilsoo Byun)

- rgw multisite: fixes for concurrent version creation ([pr#32057](#), Or Friedmann, Casey Bodley)
- rgw: prevent bucket reshards scheduling if bucket is resharding ([pr#31298](#), J. Eric Ivancich)
- rgw/pubsub: fix records/event json format to match documentation ([pr#32221](#), Yuval Lifshitz)
- rgw: radosgw-admin: sync status displays id of shard furthest behind ([pr#32818](#), Casey Bodley)
- rgw: return error if lock log shard fails ([pr#32825](#), zhangshaowen)
- rgw/rgw\_rest\_conn.h: fix build with clang ([pr#32489](#), Bernd Zeimetz)
- rgw: Select the std::bitset to resolve ambiguity ([pr#32504](#), Willem Jan Withagen)
- rgw: support radosgw-admin zone/zonegroup placement get command ([pr#32835](#), jiahuieng)
- rgw: the http response code of delete bucket should not be 204-no-content ([pr#32833](#), Chang Liu)
- rgw: update s3-test download code for s3-test tasks ([pr#32229](#), Ali Maredia)
- rgw: update the hash source for multipart entries during resharding ([pr#33183](#), dongdong tao)
- rgw: url encode common prefixes for List Objects response ([pr#32058](#), Abhishek Lekshmanan)
- rgw: when resharding store progress json ([pr#31683](#), Mark Kogan, Mark Nelson)
- selinux: Allow ceph to read udev db ([pr#32259](#), Boris Ranto)

## v14.2.7 Nautilus

---

This is the seventh update to the Ceph Nautilus release series. This is a hotfix release primarily fixing a couple of security issues. We recommend that all users upgrade to this release.

## Notable Changes

---

- CVE-2020-1699: Fixed a path traversal flaw in Ceph dashboard that could allow for potential information disclosure (Ernesto Puerta)
- CVE-2020-1700: Fixed a flaw in RGW beast frontend that could lead to denial of service from an unauthenticated client (Or Friedmann)

## v14.2.6 Nautilus

This is the sixth update to the Ceph Nautilus release series. This is a hotfix release primarily fixing a regression introduced in v14.2.5, all nautilus users are advised to upgrade to this release.

## Notable Changes

- This release fixes a `ceph-mgr` bug that caused mgr becoming unresponsive on larger clusters [issue#43364](#) ([pr#32466](#), David Zafman, Neha Ojha)

## v14.2.5 Nautilus

This is the fifth release of the Ceph Nautilus release series. Among the many notable changes, this release fixes a critical BlueStore bug that was introduced in 14.2.3. All Nautilus users are advised to upgrade to this release.

## Notable Changes

Critical fix:

- This release fixes a [critical BlueStore bug](#) introduced in 14.2.3 (and also present in 14.2.4) that can lead to data corruption when a separate “WAL” device is used.

New health warnings:

- Ceph will now issue health warnings if daemons have recently crashed. Ceph has been collecting crash reports since the initial Nautilus release, but the health alerts are new. To view new crashes (or all crashes, if you’ve just upgraded):

```
1. ceph crash ls-new
```

To acknowledge a particular crash (or all crashes) and silence the health warning:

```
1. ceph crash archive <crash-id>
2. ceph crash archive-all
```

- Ceph will issue a health warning if a RADOS pool’s `size` is set to 1 or, in other words, if the pool is configured with no redundancy. Ceph will stop issuing the warning if the pool size is set to the minimum recommended value:

```
1. ceph osd pool set <pool-name> size <num-replicas>
```

The warning can be silenced with:

```
1. ceph config set global mon_warn_on_pool_no_redundancy false
```

- A health warning is now generated if the average osd heartbeat ping time exceeds a configurable threshold for any of the intervals computed. The OSD computes 1 minute, 5 minute and 15 minute intervals with average, minimum and maximum values. New configuration option `mon_warn_on_slow_ping_ratio` specifies a percentage of `osd_heartbeat_grace` to determine the threshold. A value of zero disables the warning. New configuration option `mon_warn_on_slow_ping_time` specified in milliseconds over-rides the computed value, causes a warning when OSD heartbeat pings take longer than the specified amount. A new admin command, `ceph daemon mgr.# dump_osd_network [threshold]`, will list all connections with a ping time longer than the specified threshold or value determined by the config options, for the average for any of the 3 intervals. Another new admin command, `ceph daemon osd.# dump_osd_network [threshold]`, will do the same but only including heartbeats initiated by the specified OSD.

Changes in the telemetry module:

- The telemetry module now reports more information.

First, there is a new ‘device’ channel, enabled by default, that will report anonymized hard disk and SSD health metrics to `telemetry.ceph.com` in order to build and improve device failure prediction algorithms. If you are not comfortable sharing device metrics, you can disable that channel first before re-opting-in:

```
1. ceph config set mgr mgr/telemetry/channel_device false
```

Second, we now report more information about CephFS file systems, including:

- how many MDS daemons (in total and per file system)
- which features are (or have been) enabled
- how many data pools
- approximate file system age (year + month of creation)
- how many files, bytes, and snapshots
- how much metadata is being cached

We have also added:

- which Ceph release the monitors are running
- whether msgr v1 or v2 addresses are used for the monitors
- whether IPv4 or IPv6 addresses are used for the monitors
- whether RADOS cache tiering is enabled (and which mode)
- whether pools are replicated or erasure coded, and which erasure code profile plugin and parameters are in use
- how many hosts are in the cluster, and how many hosts have each type of daemon
- whether a separate OSD cluster network is being used
- how many RBD pools and images are in the cluster, and how many pools have RBD mirroring enabled
- how many RGW daemons, zones, and zonegroups are present; which RGW frontends are in use
- aggregate stats about the CRUSH map, like which algorithms are used, how big buckets are, how many rules are defined, and what tunables are in use

If you had telemetry enabled, you will need to re-opt-in with:

1. `ceph telemetry on`

You can view exactly what information will be reported first with:

1. `ceph telemetry show # see everything`
2. `ceph telemetry show basic # basic cluster info (including all of the new info)`

OSD:

- A new OSD daemon command, ‘dump\_recovery\_reservations’, reveals the recovery locks held (in\_progress) and waiting in priority queues.
- Another new OSD daemon command, ‘dump\_scrub\_reservations’, reveals the scrub reservations that are held for local (primary) and remote (replica) PGs.

RGW:

- RGW now supports S3 Object Lock set of APIs allowing for a WORM model for storing objects. 6 new APIs have been added put/get bucket object lock, put/get object retention, put/get object legal hold.
- RGW now supports List Objects V2

## Changelog

- bluestore/KernelDevice: fix RW\_IO\_MAX constant ([pr#31397](#), Sage Weil)
- bluestore: Don’t forget sub\_kv\_submitted\_waiters ([pr#30048](#), Jianpeng Ma)

- bluestore: apply garbage collection against excessive blob count growth ([pr#30144](#), Igor Fedotov)
- bluestore: apply shared\_alloc\_size to shared device with log level change ([pr#30229](#), Vikhyat Umrao, Sage Weil, Igor Fedotov, Neha Ojha)
- bluestore: consolidate extents from the same device only ([pr#31644](#), Igor Fedotov)
- bluestore: fix improper setting of STATE\_KV\_SUBMITTED ([pr#30755](#), Igor Fedotov)
- bluestore: shallow fsck mode and legacy statfs auto repair ([pr#30685](#), Sage Weil, Igor Fedotov)
- bluestore: tool to check fragmentation ([pr#29949](#), Adam Kupczyk)
- build/ops: admin/build-doc: use python3 ([pr#30664](#), Kefu Chai)
- build/ops: backport endian fixes ([issue#40114](#), [pr#30697](#), Ulrich Weigand, Jeff Layton)
- build/ops: cmake,rgw: IBM Z build fixes ([pr#30696](#), Ulrich Weigand)
- build/ops: cmake/BuildDPDK: ignore gcc8/9 warnings ([pr#30360](#), Yuval Lifshitz)
- build/ops: cmake: Allow cephfs and ceph-mds to be build when building on FreeBSD ([pr#31011](#), Willem Jan Withagen)
- build/ops: cmake: enforce C++17 instead of relying on cmake-compile-features ([pr#30283](#), Kefu Chai)
- build/ops: fix build fail related to PYTHON\_EXECUTABLE variable ([pr#30261](#), Ilsoo Byun)
- build/ops: hidden corei7 requirement in binary packages ([pr#29772](#), Kefu Chai)
- build/ops: install-deps.sh: add EPEL repo for non-x86\_64 archs as well ([pr#30601](#), Kefu Chai, Nathan Cutler)
- build/ops: install-deps.sh: install python\*-devel for python\*rpm-macros ([pr#30322](#), Kefu Chai)
- build/ops: install-deps: do not install if rpm already installed and ceph.spec.in: s/pkgversion/version\_nodots/ ([pr#30708](#), Jeff Layton, Kefu Chai)
- build/ops: make patch build dependency explicit ([issue#40175](#), [pr#30046](#), Nathan Cutler)
- build/ops: python3-cephfs should provide python36-cephfs ([pr#30983](#), Kefu Chai)
- build/ops: rpm: always build ceph-test package ([pr#30049](#), Nathan Cutler)
- build/ops: rpm: fdupes in SUSE builds to conform with packaging guidelines

([issue#40973](#), [pr#29784](#), Nathan Cutler)

- build/ops: rpm: make librados2, libcephfs2 own (create) /etc/ceph ([pr#31125](#), Nathan Cutler)
- build/ops: rpm: put librgw lttng S0s in the librgw-devel package ([issue#40975](#), [pr#29785](#), Nathan Cutler)
- build/ops: seastar,dmclock: use CXX\_FLAGS from parent project ([pr#30114](#), Kefu Chai)
- build/ops: use gcc-8 ([issue#38892](#), [pr#30089](#), Kefu Chai)
- tools: ceph-objectstore-tool: update-mon-db: do not fail if incmap is missing ([pr#30740](#), Kefu Chai)
- ceph-volume: PVolumes.filter shouldn't purge itself ([pr#30805](#), Rishabh Dave)
- ceph-volume: VolumeGroups.filter shouldn't purge itself ([pr#30807](#), Rishabh Dave)
- ceph-volume: add Ceph's device id to inventory ([pr#31210](#), Sebastian Wagner)
- ceph-volume: allow to skip restorecon calls ([pr#31555](#), Alfredo Deza)
- ceph-volume: api/lvm: check if list of LVs is empty ([pr#31228](#), Rishabh Dave)
- ceph-volume: check if we run in an selinux environment ([pr#31812](#), Jan Fajerski)
- ceph-volume: do not fail when trying to remove crypt mapper ([pr#30554](#), Guillaume Abrioux)
- ceph-volume: fix stderr failure to decode/encode when redirected ([pr#30300](#), Alfredo Deza)
- ceph-volume: fix warnings raised by pytest ([pr#30676](#), Rishabh Dave)
- ceph-volume: lvm list is O(n^2) ([pr#30093](#), Rishabh Dave)
- ceph-volume: lvm.zap fix cleanup for db partitions ([issue#40664](#), [pr#30304](#), Dominik Csapak)
- ceph-volume: mokeypatch calls to lvm related binaries ([pr#31405](#), Jan Fajerski)
- ceph-volume: pre-install python-apt and its variants before test runs ([pr#30294](#), Alfredo Deza)
- ceph-volume: rearrange api/lvm.py ([pr#31408](#), Rishabh Dave)
- ceph-volume: systemd fix typo in log message ([pr#30520](#), Manu Zurmühl)
- ceph-volume: use the OSD identifier when reporting success ([pr#29769](#), Alfredo Deza)

- ceph-volume: zap always skips block.db, leaves them around ([issue#40664](#), [pr#30307](#), Alfredo Deza)
- tools: ceph.in: do not preload ASan unless necessary ([pr#31676](#), Kefu Chai)
- build/ops: ceph.spec.in: reserve 2500MB per build job ([pr#30370](#), Dan van der Ster)
- tools: ceph\_volume\_client: convert string to bytes object ([issue#39405](#), [issue#40369](#), [issue#39510](#), [issue#40800](#), [issue#40460](#), [pr#30030](#), Rishabh Dave)
- cephfs-shell: Convert paths type from string to bytes ([pr#30057](#), Varsha Rao)
- cephfs: Allow mount.ceph to get mount info from ceph configs and keyrings ([pr#30521](#), Jeff Layton)
- cephfs: avoid map been inserted by mistake ([pr#29878](#), XiaoGuoDong2019)
- cephfs: client: more precise CEPH\_CLIENT\_CAPS\_PENDING\_CAPSNAP ([pr#30032](#), "Yan, Zheng")
- cephfs: client: nfs-ganesha with cephfs client, removing dir reports not empty ([issue#40746](#), [pr#30442](#), Peng Xie)
- cephfs: client: return -eio when sync file which unsafe reqs have been dropped ([issue#40877](#), [pr#30043](#), simon gao)
- cephfs: fix a memory leak ([pr#29879](#), XiaoGuoDong2019)
- cephfs: mds: Fix duplicate client entries in eviction list ([pr#30951](#), Sidharth Anupkrishnan)
- cephfs: mds: cleanup truncating inodes when standby replay mds trim log segments ([pr#29591](#), "Yan, Zheng")
- cephfs: mds: delay exporting directory whose pin value exceeds max rank id ([issue#40603](#), [pr#29938](#), Zhi Zhang)
- cephfs: mds: evict an unresponsive client only when another client wants its caps ([pr#30031](#), Rishabh Dave)
- cephfs: mds: fix InoTable::force\_consume\_to() ([pr#30041](#), "Yan, Zheng")
- cephfs: mds: fix infinite loop in Locker::file\_update\_finish ([pr#31079](#), "Yan, Zheng")
- cephfs: mds: make MDSIOContextBase delete itself when shutting down ([pr#30418](#), Xuehan Xu)
- cephfs: mds: trim cache on regular schedule ([pr#30040](#), Patrick Donnelly)
- cephfs: mds: wake up lock waiters after forcibly changing lock state

- ([issue#39987](#), [pr#30508](#), "Yan, Zheng")
- cephfs: mount.ceph: properly handle -o strictatime ([pr#30039](#), Jeff Layton)
  - cephfs: qa: ignore expected MDS\_CLIENT\_LATE\_RELEASE warning ([issue#40968](#), [pr#29811](#), Patrick Donnelly)
  - cephfs: qa: wait for MDS to come back after removing it ([issue#40967](#), [pr#29832](#), Patrick Donnelly)
  - cephfs: tests: power off still resulted in client sending session close ([issue#37681](#), [pr#29983](#), Patrick Donnelly)
  - common/ceph\_context: avoid unnecessary wait during service thread shutdown ([pr#31097](#), Jason Dillaman)
  - common/config\_proxy: hold lock while accessing mutable container ([pr#30661](#), Jason Dillaman)
  - common: fix typo in rgw\_user\_max\_buckets option long description ([pr#31605](#), Alfonso Martínez)
  - core/osd: do not trust partially simplified pg\_upmap\_item ([issue#42052](#), [pr#30899](#), xie xingguo)
  - core: Health warnings on long network ping times ([issue#40640](#), [pr#30195](#), David Zafman)
  - core: If the nodeep-scrub/noscrub flags are set in pools instead of global cluster. List the pool names in the ceph status ([issue#38029](#), [pr#29991](#), Mohamad Gebai)
  - core: Improve health status for backfill\_toofull and recovery\_toofull and fix backfill\_toofull seen on cluster where the most full OSD is at 1% ([pr#29999](#), David Zafman)
  - core: Make dumping of reservation info congruent between scrub and recovery ([pr#31444](#), David Zafman)
  - core: Revert “rocksdb: enable rocksdb\_rmrang=true by default” ([pr#31612](#), Neha Ojha)
  - core: filestore pre-split may not split enough directories ([issue#39390](#), [pr#29988](#), Jeegn Chen)
  - core: kv/RocksDBStore: tell rocksdb to set mode to 0600, not 0644 ([pr#31031](#), Sage Weil)
  - core: mon/MonClient: ENXIO when sending command to down mon ([pr#31037](#), Sage Weil, Greg Farnum)

- core: mon/MonCommands: “smart” only needs read permission ([pr#31111](#), Kefu Chai)
- core: mon/MonMap: encode (more) valid compat monmap when we have v2-only addrs ([pr#31658](#), Sage Weil)
- core: mon/Monitor.cc: fix condition that checks for unrecognized auth mode ([pr#31038](#), Neha Ojha)
- core: mon/OSDMonitor: Use generic priority cache tuner for mon caches ([pr#30419](#), Sridhar Seshasayee, Kefu Chai, Mykola Golub, Mark Nelson)
- core: mon/OSDMonitor: add check for crush rule size in pool set size command ([pr#30941](#), Vikhyat Umrao)
- core: mon/OSDMonitor: trim not-longer-exist failure reporters ([pr#30904](#), NancySu05)
- core: mon/PGMap: fix incorrect pg\_pool\_sum when delete pool ([pr#31704](#), luo rixin)
- core: mon: C\_AckMarkedDown has not handled the Callback Arguments ([pr#29997](#), NancySu05)
- core: mon: ensure prepare\_failure() marks no\_reply on op ([pr#30480](#), Joao Eduardo Luis)
- core: mon: show pool id in pool ls command ([issue#40287](#), [pr#30486](#), Chang Liu)
- core: msg,mon/MonClient: fix auth for clients without CEPHX\_V2 feature ([pr#30524](#), Sage Weil)
- core: msg/auth: handle decode errors instead of throwing exceptions ([pr#31099](#), Sage Weil)
- core: msg/simple: reset in\_seq\_acked to zero when session is reset ([pr#29592](#), Xiangyang Yu)
- core: os/bluestore: fix objectstore\_blackhole read-after-write ([pr#31019](#), Sage Weil)
- core: osd/OSDCap: Check for empty namespace ([issue#40835](#), [pr#29998](#), Brad Hubbard)
- core: mon/OSDMonitor: make memory autotune disable itself if no rocksdb ([pr#32045](#), Sage Weil)
- core: osd/PG: Add PG to large omap log message ([pr#30923](#), Brad Hubbard)
- core: osd/PGLog: persist num\_objects\_missing for replicas when peering is done ([pr#31077](#), xie xingguo)
- core: osd/PeeringState: do not complain about past\_intervals constrained by oldest epoch ([pr#30000](#), Sage Weil)

- core: osd/PeeringState: fix wrong history of merge target ([pr#30280](#), xie xingguo)
- core: osd/PeeringState: recover\_got - add special handler for empty log and improvements to standalone tests ([pr#30528](#), Sage Weil, David Zafman, xie xingguo)
- core: osd/PrimaryLogPG: Avoid accessing destroyed references in finish\_degr... ([pr#29994](#), Tao Ning)
- core: osd/PrimaryLogPG: update oi.size on write op implicitly truncating ob... ([pr#30278](#), xie xingguo)
- core: osd/ReplicatedBackend: check against empty data\_included before enabling crc ([pr#29716](#), xie xingguo)
- core: osd/osd\_types: fix {omap,hitset\_bytes}\_stats\_invalid handling on spli... ([pr#30643](#), Sage Weil)
- core: osd: Better error message when OSD count is less than osd\_pool\_default\_size ([issue#38617](#), [pr#29992](#), Kefu Chai, Sage Weil, zjh)
- core: osd: Remove unused osdmap flags full, nearfull from output ([pr#30900](#), David Zafman)
- core: osd: add log information to record the cause of do\_osd\_ops failure ([pr#30546](#), NancySu05)
- core: osd: clear PG\_STATE\_CLEAN when repair object ([pr#30050](#), Zengran Zhang)
- core: osd: fix possible crash on sending dynamic perf stats report ([pr#30648](#), Mykola Golub)
- core: osd: merge replica log on primary need according to replica log's crt ([pr#30051](#), Zengran Zhang)
- core: osd: prime splits/merges for any potential fabricated split/merge par... ([issue#38483](#), [pr#30371](#), xie xingguo)
- core: osd: release backoffs during merge ([pr#31822](#), Sage Weil)
- core: osd: rollforward may need to mark pglog dirty ([issue#40403](#), [pr#31034](#), Zengran Zhang)
- core: osd: scrub error on big objects; make bluestore refuse to start on big objects ([pr#30783](#), David Zafman, Sage Weil)
- core: osd: support osd\_repair\_during\_recovery ([issue#40620](#), [pr#29748](#), Jeegn Chen)
- core: pool\_stat.dump() - value of num\_store\_stats is wrong ([issue#39340](#), [pr#29946](#), xie xingguo)
- doc/ceph-kvstore-tool: add description for 'stats' command ([pr#30245](#), Josh Durgin, Adam Kupczyk)

- doc/mgr/telemetry: update default interval ([pr#31009](#), Tim Serong)
- doc/rbd: s/guess/xml/ for codeblock lexer ([pr#31074](#), Kefu Chai)
- doc: Fix rbd namespace documentation ([pr#29731](#), Ricardo Marques)
- doc: cephfs: add section on fsync error reporting to posix.rst ([issue#24641](#), [pr#30025](#), Jeff Layton)
- doc: default values for mon\_health\_to\_clog\_\* were flipped ([pr#30003](#), James McClune)
- doc: fix urls in posix.rst ([pr#30686](#), Jos Collin)
- doc: max\_misplaced option was renamed in Nautilus ([pr#30649](#), Nathan Fish)
- doc: pg\_num should always be a power of two ([pr#30004](#), Lars Marowsky-Bree, Kai Wagner)
- doc: update bluestore cache settings and clarify data fraction ([issue#39522](#), [pr#31259](#), Jan Fajerski)
- mgr/ActivePyModules: behave if a module queries a devid that does not exist ([pr#31411](#), Sage Weil)
- mgr/BaseMgrStandbyModule: drop GIL in ceph\_get\_module\_option() ([pr#30773](#), Kefu Chai)
- mgr/balancer: python3 compatibility issue ([pr#31012](#), Mykola Golub)
- mgr/crash: backport archive feature, health alerts ([pr#30851](#), Sage Weil)
- mgr/crash: try client.crash[.host] before client.admin; add mon profile ([issue#40781](#), [pr#30844](#), Sage Weil, Dan Mick)
- mgr/dashboard: Add transifex-i18ntool ([pr#31160](#), Sebastian Krah)
- mgr/dashboard: Allow disabling redirection on standby dashboards ([issue#41813](#), [pr#30382](#), Volker Theile)
- mgr/dashboard: Configuring an URL prefix does not work as expected ([pr#31375](#), Volker Theile)
- mgr/dashbaord: Fix calculation of PG status percentage ([issue#41809](#), [pr#30394](#), Tiago Melo)
- mgr/dashboard: Fix CephFS chart ([pr#30691](#), Stephan Müller)
- mgr/dashboard: Fix grafana dashboards ([pr#31733](#), Radu Toader)
- mgr/dashboard: Improve position of MDS chart tooltip ([pr#31565](#), Tiago Melo)

- mgr/dashboard: Provide the name of the object being deleted ([pr#31263](#), Ricardo Marques)
- mgr/dashboard: RBD tests must use pools with power-of-two pg\_num ([pr#31522](#), Ricardo Marques)
- mgr/dashboard: Set RO as the default access\_type for RGW NFS exports ([pr#30516](#), Tiago Melo)
- mgr/dashboard: Wait for breadcrumb text is present in e2e tests ([pr#31576](#), Volker Theile)
- mgr/dashboard: access\_control: add grafana scope read access to \*-manager roles ([pr#30259](#), Ricardo Dias)
- mgr/dashboard: do not log tokens ([pr#31413](#), Kefu Chai)
- mgr/dashboard: do not show non-pool data in pool details ([pr#31516](#), Alfonso Martínez)
- mgr/dashboard: edit/clone/copy rbd image after its data is received ([pr#31349](#), Alfonso Martínez)
- mgr/dashboard: internationalization support with AOT enabled ([pr#30910](#), Ricardo Dias, Tiago Melo)
- mgr/dashboard: run-backend-api-tests.sh improvements ([pr#29487](#), Alfonso Martínez, Kefu Chai)
- mgr/dashboard: tasks: only unblock controller thread after TaskManager thread ([pr#31526](#), Ricardo Dias)
- mgr/devicehealth: do not scrape mon devices ([pr#31446](#), Sage Weil)
- mgr/devicehealth: import \_strptime directly ([pr#32082](#), Sage Weil)
- mgr/k8sevents: Initial ceph -> k8s events integration ([pr#30215](#), Paul Cuzner, Sebastian Wagner)
- mgr/pg\_autoscaler: fix pool\_logical\_used ([pr#31100](#), Ansgar Jazdzewski)
- mgr/pg\_autoscaler: fix race with pool deletion ([pr#30008](#), Sage Weil)
- mgr/prometheus: Cast collect\_timeout (scrape\_interval) to float ([pr#30007](#), Ben Meekhof)
- mgr/prometheus: Fix KeyError in get\_mgr\_status ([pr#30774](#), Sebastian Wagner)
- mgr/rbd\_support: module.py:1088: error: Name 'image\_spec' is not defined ([pr#29978](#), Jason Dillaman)
- mgr/restful: requests api adds support multiple commands ([pr#31334](#), Duncan

Chiang)

- mgr/telemetry: backport a ton of stuff ([pr#30849](#), alfonsomthd, Kefu Chai, Sage Weil, Dan Mick)
- mgr/volumes: fix incorrect snapshot path creation ([pr#31076](#), Ramana Raja)
- mgr/volumes: handle exceptions in purge thread with retry ([issue#41218](#), [pr#30455](#), Venky Shankar)
- mgr/volumes: list FS subvolumes, subvolume groups, and their snapshots ([pr#30827](#), Jos Collin)
- mgr/volumes: minor fixes ([pr#29926](#), Venky Shankar, Jos Collin, Ramana Raja)
- mgr/volumes: protection for “fs volume rm” command ([pr#30768](#), Jos Collin, Ramana Raja)
- mgr/zabbix: Fix typo in key name for PGs in backfill\_wait state ([issue#39666](#), [pr#30006](#), Wido den Hollander)
- mgr/zabbix: encode string for Python 3 compatibility ([pr#30016](#), Nathan Cutler)
- mgr/{dashboard,prometheus}: return FQDN instead of ‘0.0.0.0’ ([pr#31482](#), Patrick Seidensal)
- mgr: Release GIL before calling OSDMap::calc\_pg\_upmaps() ([pr#31682](#), David Zafman, Shyukri Shyukriev)
- mgr: Unable to reset / unset module options ([issue#40779](#), [pr#29550](#), Sebastian Wagner)
- mgr: do not reset reported if a new metric is not collected ([pr#30390](#), Ilsoo Byun)
- mgr: fix weird health-alert daemon key ([pr#31039](#), xie xingguo)
- mgr: set hostname in DeviceState::set\_metadata() ([pr#30624](#), Kefu Chai)
- pybind/cephfs: Modification to error message ([pr#30026](#), Varsha Rao)
- pybind/rados: fix set\_omap() crash on py3 ([pr#30622](#), Sage Weil)
- pybind/rbd: deprecate parent\_info ([pr#30818](#), Ricardo Marques)
- rbd: rbd-mirror: cannot restore deferred deletion mirrored images ([pr#30825](#), Jason Dillaman, Mykola Golub)
- rbd: rbd-mirror: don’t overwrite status error returned by replay ([pr#29870](#), Mykola Golub)
- rbd: rbd-mirror: ignore errors relating to parsing the cluster config file

([pr#30116](#), Jason Dillaman)

- rbd: rbd-mirror: simplify peer bootstrapping ([pr#30821](#), Jason Dillaman)
- rbd: rbd-nbd: add netlink support and nl resize ([pr#30532](#), Mike Christie)
- rbd: cls/rbd: sanitize entity instance messenger version type ([pr#30822](#), Jason Dillaman)
- rbd: cls/rbd: sanitize the mirror image status peer address after reading from disk ([pr#31833](#), Jason Dillaman)
- rbd: krbd: avoid udev netlink socket overrun and retry on transient errors from udev\_enumerate\_scan\_devices() ([pr#31075](#), Ilya Dryomov, Adam C. Emerson)
- rbd: librbd: always try to acquire exclusive lock when removing image ([pr#29869](#), Mykola Golub)
- rbd: librbd: behave more gracefully when data pool removed ([pr#30824](#), Mykola Golub)
- rbd: librbd: v1 clones are restricted to the same namespace ([pr#30823](#), Jason Dillaman)
- mgr/restful: Query nodes\_by\_id for items ([pr#31261](#), Boris Ranto)
- rgw/amqp: fix race condition in AMQP unit test ([pr#30889](#), Yuval Lifshitz)
- rgw/amqp: remove flaky amqp test ([pr#31628](#), Yuval Lifshitz)
- rgw/pubsub: backport notifications and pubsub ([pr#30579](#), Yuval Lifshitz)
- rgw/rgw\_op: Remove get\_val from hotpath via legacy options ([pr#30160](#), Mark Nelson)
- rgw: Potential crash in putbj ([pr#29898](#), Adam C. Emerson)
- rgw: Put User Policy is sensitive to whitespace ([pr#29970](#), Abhishek Lekshmanan)
- rgw: RGWCoroutine::call(nullptr) sets retcode=0 ([pr#30248](#), Casey Bodley)
- rgw: Swift metadata dropped after S3 bucket versioning enabled ([pr#29961](#), Marcus Watts)
- rgw: add S3 object lock feature to support object worm ([pr#29905](#), Chang Liu, Casey Bodley, zhang Shaowen)
- rgw: add minssing admin property when sync user info ([pr#30680](#), zhang Shaowen)
- rgw: beast frontend throws an exception when running out of FDs ([pr#29963](#), Yuval Lifshitz)

- rgw: data/bilogs are trimmed when no peers are reading them ([issue#39487](#), [pr#30999](#), Casey Bodley)
- rgw: datalog/mdlog trim commands loop until done ([pr#30869](#), Casey Bodley)
- rgw: dns name is not case sensitive ([issue#40995](#), [pr#29971](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: fix a bug that lifecycle expiraton generates delete marker continuously ([issue#40393](#), [pr#30037](#), zhang Shaowen)
- rgw: fix cls\_bucket\_list\_unordered() partial results ([pr#30252](#), Mark Kogan)
- rgw: fix data sync start delay if remote haven't init data\_log ([pr#30509](#), Tianshan Qu)
- rgw: fix default storage class for get\_compression\_type ([pr#31026](#), Casey Bodley)
- rgw: fix drain handles error when deleting bucket with bypass-gc option ([pr#29956](#), dongdong tao)
- rgw: fix list bucket with delimiter wrongly skip some special keys ([issue#40905](#), [pr#30068](#), Tianshan Qu)
- rgw: fix memory growth while deleteing objects with ([pr#30472](#), Mark Kogan)
- rgw: fix the bug of rgw not doing necessary checking to website configuration ([issue#40678](#), [pr#30325](#), Enming Zhang)
- rgw: fixed "unrecognized arg" error when using "radosgw-admin zone rm" ([pr#30247](#), Hongang Chen)
- rgw: housekeeping reset stats ([pr#29803](#), J. Eric Ivancich)
- rgw: increase beast parse buffer size to 64k ([pr#30437](#), Casey Bodley)
- rgw: ldap auth: S3 auth failure should return InvalidAccessKeyId ([pr#30651](#), Matt Benjamin)
- rgw: lifecycle days may be 0 ([pr#31073](#), Matt Benjamin)
- rgw: lifecycle transitions on non existent placement targets ([pr#29955](#), Abhishek Lekshmanan)
- rgw: list objects version 2 ([pr#29849](#), Albin Antony, zhang Shaowen)
- rgw: multisite: radosgw-admin bucket sync status incorrectly reports "caught up" during full sync ([issue#40806](#), [pr#29974](#), Casey Bodley)
- rgw: potential realm watch lost ([issue#40991](#), [pr#29972](#), Tianshan Qu)
- rgw: protect AioResultList by a lock to avoid race condition ([pr#30746](#), Ilsoo)

Byun)

- rgw: radosgw-admin: add -uid check in bucket list command ([pr#30604](#), Vikhyat Umrao)
- rgw: returns one byte more data than the requested range from the SLO object ([pr#29960](#), Andrey Groshev)
- rgw: rgw-admin: search for user by access key ([pr#29959](#), Matt Benjamin)
- rgw: rgw-log issues the wrong message when decompression fails ([pr#29965](#), Han Fengzhe)
- rgw: rgw\_file: directory enumeration can be accelerated 1-2 orders of magnitude taking stats from bucket index Part I (stats from S3/Swift only) ([issue#40456](#), [pr#29954](#), Matt Benjamin)
- rgw: rgw\_file: readdir: do not construct markers w/leading '/' ([pr#29969](#), Matt Benjamin)
- rgw: silence warning "control reaches end of non-void function" ([issue#40747](#), [pr#31742](#), Jos Collin)
- rgw: sync with elastic search v7 ([pr#31027](#), Chang Liu)
- rgw: use explicit `to_string()` overload for `boost::string_ref` ([issue#39611](#), [pr#31650](#), Casey Bodley, Ulrich Weigand)
- rgw: when using radosgw-admin to list bucket, can set `-max-entries` excessively high ([pr#29777](#), J. Eric Ivancich)
- tests: "CMake Error" in `test_envlibrados_for_rocksdb.sh` ([pr#29979](#), Kefu Chai)
- tests: Get libcephfs and cephfs to compile with FreeBSD ([pr#31136](#), Willem Jan Withagen)
- tests: add debugging failed osd-release setting ([pr#31040](#), Patrick Donnelly)
- tests: cephfs: fix malformed qa suite config ([pr#30038](#), Patrick Donnelly)
- tests: cls\_rbd/test\_cls\_rbd: update `TestClsRbd.sparsify` ([pr#30354](#), Kefu Chai)
- tests: cls\_rbd: removed mirror peer pool test cases ([pr#30948](#), Jason Dillaman)
- tests: enable dashboard tests to be run with "-suite rados/dashboard" ([pr#31248](#), Nathan Cutler)
- tests: librbd: set nbd timeout due to newer kernels defaulting it on ([pr#30423](#), Jason Dillaman)
- tests: qa/suites/krbd: run unmap subsuite with msgr1 only ([pr#31290](#), Ilya Dryomov)

- tests: qa/tasks/cbt: run stop-all.sh while shutting down ([pr#31304](#), Sage Weil)
- tests: qa/tasks/ceph.conf.template: increase mon tell retries ([pr#31641](#), Sage Weil)
- tests: qa/workunits/rbd: stress test rbd mirror pool status -verbose ([pr#29871](#), Mykola Golub)
- tests: qa: avoid page cache for krbd discard round off tests ([pr#30464](#), Ilya Dryomov)
- tests: qa: sleep briefly after resetting kclient ([pr#29750](#), Patrick Donnelly)
- tests: rados/mgr/tasks/module\_selftest: whitelist mgr client getting blacklisted ([issue#40867](#), [pr#29649](#), Sage Weil)
- tests: test\_librados\_build.sh: grab from nautilus branch in nautilus ([pr#31604](#), Nathan Cutler)
- tests: valgrind: UninitCondition in ceph::crypto::onwire::AES128GCM\_OnWireRxHandler::authenticated\_decrypt\_update\_final() ([issue#38827](#), [pr#29928](#), Radoslaw Zarzynski)
- tools/rados: add -pgid in help ([pr#30607](#), Vikhyat Umrao)
- tools/rados: call pool\_lookup() after rados is connected ([pr#30605](#), Vikhyat Umrao)
- tools/rbd-ggate: close log before running postfork ([pr#30120](#), Willem Jan Withagen)
- tools: ceph-backport.sh: add deprecation warning ([pr#30748](#), Nathan Cutler)
- tools: ceph-objectstore-tool can't remove head with bad snapset ([pr#30080](#), David Zafman)

## v14.2.4 Nautilus

---

This is the fourth release in the Ceph Nautilus stable release series. Its sole purpose is to fix a regression that found its way into the previous release.

## Notable Changes

---

- The ceph-volume in Nautilus v14.2.3 was found to contain a serious regression, described in <https://tracker.ceph.com/issues/41660>, which prevented deployment tools like ceph-ansible, DeepSea, Rook, etc. from deploying/removing OSDs.

## Changelog

---

- ceph-volume: fix stderr failure to decode/encode when redirected ([pr#30300](#), Alfredo Deza)

## v14.2.3 Nautilus

---

This is the third bug fix release of Ceph Nautilus release series. We recommend all Nautilus users upgrade to this release. For upgrading from older releases of ceph, general guidelines for upgrade to nautilus must be followed [Upgrading from Mimic or Luminous](#).

## Notable Changes

---

- CVE-2019-10222 - Fixed a denial of service vulnerability where an unauthenticated client of Ceph Object Gateway could trigger a crash from an uncaught exception
- Nautilus-based librbd clients can now open images on Jewel clusters.
- The RGW num\_rados\_handles has been removed. If you were using a value of num\_rados\_handles greater than 1, multiply your current objecter\_inflight\_ops and objecter\_inflight\_op\_bytes parameters by the old num\_rados\_handles to get the same throttle behavior.
- The secure mode of Messenger v2 protocol is no longer experimental with this release. This mode is now the preferred mode of connection for monitors.
- “osd\_deep\_scrub\_large\_omap\_object\_key\_threshold” has been lowered to detect an object with large number of omap keys more easily.
- The Ceph Dashboard now supports silencing Prometheus alert notifications.

## Changelog

---

- bluestore: 50-100% iops lost due to bluefs\_preeextend\_wal\_files = false ([issue#38559](#), [pr#28573](#), Vitaliy Filippov)
- bluestore: add slow op detection for collection\_listing ([pr#29227](#), Igor Fedotov)
- bluestore: avoid length overflow in extents returned by Stupid Allocator ([issue#40703](#), [pr#29023](#), Igor Fedotov)
- bluestore/bluefs\_types: consolidate contiguous extents ([pr#28862](#), Sage Weil)
- bluestore/bluestore-tool: minor fixes around migrate ([pr#28893](#), Igor Fedotov)
- bluestore: create the tail when first set FLAG\_OMAP ([issue#36482](#), [pr#28963](#), Tao Ning)
- bluestore: do not set osd\_memory\_target default from cgroup limit ([pr#29745](#), Sage Weil)

Weil)

- bluestore: fix >2GB bluefs writes ([pr#28966](#), kungf, Sage Weil)
- bluestore: load OSD all compression settings unconditionally ([issue#40480](#), [pr#28892](#), Igor Fedotov)
- bluestore: more smart allocator dump when lacking space for bluefs ([issue#40623](#), [pr#28891](#), Igor Fedotov)
- bluestore: Set concurrent max\_background\_compactions in rocksdb to 2 ([issue#40769](#), [pr#29162](#), Mark Nelson)
- bluestore: support RocksDB prefetch in buffered read mode ([pr#28962](#), Igor Fedotov)
- build/ops: Module 'dashboard' has failed: No module named routes ([issue#24420](#), [pr#28992](#), Paul Emmerich)
- build/ops: rpm: drop SuSEfirewall2 ([issue#40738](#), [pr#29007](#), Matthias Gerstner)
- build/ops: rpm: Require ceph-grafana-dashboards ([pr#29682](#), Boris Ranto)
- cephfs: ceph-fuse: mount does not support the fallocate() ([issue#40615](#), [pr#29157](#), huanwen ren)
- cephfs: ceph\_volume\_client: d\_name needs to be converted to string before using ([issue#39406](#), [pr#28609](#), Rishabh Dave)
- cephfs: client: bump ll\_ref from int32 to uint64\_t ([pr#29186](#), Xiaoxi CHEN)
- cephfs: client: set snapdir's link count to 1 ([issue#40101](#), [pr#29343](#), "Yan, Zheng")
- cephfs: client: unlink dentry for inode with llref=0 ([issue#40960](#), [pr#29478](#), Xiaoxi CHEN)
- cephfs: getattr on snap inode stuck ([issue#40361](#), [pr#29231](#), "Yan, Zheng")
- cephfs: mds: cannot switch mds state from standby-replay to active ([issue#40213](#), [pr#29233](#), simon gao)
- cephfs: mds: cleanup unneeded client\_snap\_caps when splitting snap inode ([issue#39987](#), [pr#29344](#), "Yan, Zheng")
- cephfs-shell: name 'files' is not defined error in do\_rm() ([issue#40489](#), [pr#29158](#), Varsha Rao)
- cephfs-shell: TypeError in poutput ([issue#40679](#), [pr#29156](#), Varsha Rao)
- ceph.spec.in: Drop systemd BuildRequires in case of building for SUSE ([pr#28937](#), Dominique Leuenberger)

- ceph-volume: batch functional idempotency test fails since message is now on stderr ([pr#29689](#), Jan Fajerski)
- ceph-volume: batch gets confused when the same device is passed in two device lists ([pr#29690](#), Jan Fajerski)
- ceph-volume: does not recognize wal/db partitions created by ceph-disk ([pr#29464](#), Jan Fajerski)
- ceph-volume: [filestore,bluestore] single type strategies fail after tracking devices as sets ([pr#29702](#), Jan Fajerski)
- ceph-volume: lvm.activate: Return an error if WAL/DB devices absent ([pr#29040](#), David Casier)
- ceph-volume: missing string substitution when reporting mounts ([issue#25030](#), [pr#29260](#), Shyukri Shyukriev)
- ceph-volume: prints errors to stdout with -format json ([issue#38548](#), [pr#29506](#), Jan Fajerski)
- ceph-volume: prints log messages to stdout ([pr#29600](#), Jan Fajerski, Kefu Chai, Alfredo Deza)
- ceph-volume: run functional tests without dashboard ([pr#29694](#), Andrew Schoen)
- ceph-volume: simple functional tests drop test for lvm zap ([pr#29660](#), Jan Fajerski)
- ceph-volume: tests set the noninteractive flag for Debian ([pr#29899](#), Alfredo Deza)
- ceph-volume: when 'type' file is not present activate fails ([pr#29416](#), Alfredo Deza)
- cmake: update FindBoost.cmake ([pr#29436](#), Willem Jan Withagen)
- common/config: respect POD\_MEMORY\_REQUEST \*and\* POD\_MEMORY\_LIMIT env vars ([pr#29562](#), Patrick Donnelly, Sage Weil)
- common: Keyrings created by ceph auth get are not suitable for ceph auth import ([issue#22227](#), [pr#28740](#), Kefu Chai)
- common: OutputDataSocket retakes mutex on error path ([issue#40188](#), [pr#29147](#), Casey Bodley)
- core: Better default value for osd\_snap\_trim\_sleep ([pr#29678](#), Neha Ojha)
- core: Change default for bluestore\_fsck\_on\_mount\_deep as false ([pr#29697](#), Neha Ojha)
- core: lazy omap stat collection ([pr#29188](#), Brad Hubbard)

- core: librados: move buffer free functions to inline namespace ([issue#39972](#), [pr#29244](#), Jason Dillaman)
- core: maybe\_remove\_pg\_upmap can be super inefficient for large clusters ([issue#40104](#), [pr#28756](#), xie xingguo)
- core: MDSMonitor: use stringstream instead of dout for mds repaired ([issue#40472](#), [pr#29159](#), Zhi Zhang)
- core: osd beacon sometimes has empty pg list ([issue#40377](#), [pr#29254](#), Sage Weil)
- core: s3tests-test-readwrite failed in rados run (Connection refused) ([issue#17882](#), [pr#29325](#), Casey Bodley)
- doc: Document more cache modes ([issue#14153](#), [pr#28958](#), Nathan Cutler)
- doc: fix rgw ldap username token ([pr#29455](#), Thomas Kriechbaumer)
- doc: Improved dashboard feature overview ([pr#28919](#), Lenz Grimmer)
- doc: Object Gateway multisite document read-only argument error ([issue#40458](#), [pr#29306](#), Chenjiong Deng)
- doc/rados: Correcting some typos in the clay code documentation ([pr#29191](#), Myna)
- doc/rbd: initial live-migration documentation ([issue#40486](#), [pr#29724](#), Jason Dillaman)
- doc/rgw: document use of 'realm pull' instead of 'period pull' ([issue#39655](#), [pr#29484](#), Casey Bodley)
- doc: steps to disable metadata\_heap on existing rgw zones ([issue#18174](#), [pr#28738](#), Dan van der Ster)
- doc: Update 'ceph-iscsi' min version ([pr#29444](#), Ricardo Marques)
- journal: properly advance read offset after skipping invalid range ([pr#28816](#), Mykola Golub)
- librbd: improve journal performance to match expected degredation ([issue#40072](#), [pr#29723](#), Mykola Golub, Jason Dillaman)
- librbd: properly track in-flight flush requests ([issue#40555](#), [pr#28769](#), Jason Dillaman)
- librbd: snapshot object maps can go inconsistent during copyup ([issue#39435](#), [pr#29722](#), Ilya Dryomov)
- mds: change how mds revoke stale caps ([issue#17854](#), [pr#28583](#), Rishabh Dave, "Yan, Zheng")
- mgr: Add mgr metdata to prometheus exporter module ([pr#29168](#), Paul Cuzner)

- mgr/dashboard: Add, update and remove translations ([issue#39701](#), [pr#28938](#), Sebastian Krah)
- mgr/dashboard: cephfs multimds graphs stack together ([issue#37579](#), [pr#28889](#), Kiefer Chang)
- mgr/dashboard: Changing rgw-api-host does not get effective without disable/enable dashboard mgr module ([issue#40252](#), [pr#29044](#), Ricardo Marques)
- mgr/dashboard: controllers/grafana is not Python3 compatible ([issue#40428](#), [pr#29524](#), Patrick Nawracay)
- mgr/dashboard: Dentries value of MDS daemon in Filesystems page is inconsistent with ceph fs status output ([issue#40097](#), [pr#28912](#), Kiefer Chang)
- mgr/dashboard: Display logged in information for each iSCSI client ([issue#40046](#), [pr#29045](#), Ricardo Marques)
- mgr/dashboard: Fix e2e failures caused by webdriver version ([pr#29491](#), Tiago Melo)
- mgr/dashboard: Fix npm vulnerabilities ([issue#40677](#), [pr#29102](#), Tiago Melo)
- mgr/dashboard: Fix the table mouseenter event handling test ([issue#40580](#), [pr#29354](#), Stephan Müller)
- mgr/dashboard: Interlock fast-diff and object-map ([issue#39451](#), [pr#29442](#), Patrick Nawracay)
- mgr/dashboard: notify the user about unset 'mon\_allow\_pool\_delete' flag beforehand ([issue#39533](#), [pr#28833](#), Tatjana Dehler)
- mgr/dashboard: Optimize the calculation of portal IPs ([issue#39580](#), [pr#29061](#), Ricardo Marques, Kefu Chai)
- mgr/dashboard: Pool graph/sparkline points do not display the correct values ([issue#39650](#), [pr#29352](#), Stephan Müller)
- mgr/dashboard: RGW User quota validation is not working correctly ([pr#29650](#), Volker Theile)
- mgr/dashboard: Silence Alertmanager alerts ([issue#36722](#), [pr#28968](#), Stephan Müller)
- mgr/dashboard: SSL certificate upload command throws deprecation warning ([issue#39123](#), [pr#29065](#), Ricardo Dias)
- mgr/dashboard: switch ng2-toastr to ngx-toastr ([pr#29050](#), Tiago Melo, Ernesto Puerta)
- mgr/dashboard: Upgrade to ceph-iscsi config v10 ([issue#40566](#), [pr#28974](#), Ricardo

Marques)

- mgr/diskprediction\_cloud: Service unavailable ([issue#40478](#), [pr#29454](#), Rick Chen)
- mgr/influx: module fails due to missing close() method ([issue#40174](#), [pr#29207](#), Kefu Chai)
- mgr/orchestrator: Cache and DeepSea iSCSI + NFS ([pr#29060](#), Sebastian Wagner, Tim Serong)
- mgr/rbd\_support: support scheduling long-running background operations ([issue#40621](#), [issue#40790](#), [pr#29725](#), Venky Shankar, Jason Dillaman)
- mgr: use ipv4 default when ipv6 was disabled ([issue#40023](#), [pr#29194](#), kungf)
- mgr/volumes: background purge queue for subvolumes ([issue#40036](#), [pr#29079](#), Patrick Donnelly, Venky Shankar, Kefu Chai)
- mgr/volumes: minor enhancement and bug fix ([issue#40927](#), [issue#40617](#), [pr#29490](#), Ramana Raja)
- mon: auth mon isn't loading full KeyServerData after restart ([issue#40634](#), [pr#28993](#), Sage Weil)
- mon/MgrMonitor: fix null deref when invalid formatter is specified ([pr#29566](#), Sage Weil)
- mon/OSDMonitor: allow pg\_num to increase when require\_osd\_release < N ([issue#39570](#), [pr#29671](#), Neha Ojha, Sage Weil)
- mon/OSDMonitor.cc: better error message about min\_size ([pr#29617](#), Neha Ojha)
- mon: paxos: introduce new reset\_pending\_committing\_finishers for safety ([issue#39484](#), [pr#28528](#), Greg Farnum)
- mon: set recovery priority etc on cephfs metadata pool ([pr#29275](#), Sage Weil)
- mon: take the mon lock in handle\_conf\_change ([issue#39625](#), [pr#29373](#), huangjun)
- msg/async: avoid unnecessary costly wakeups for outbound messages ([pr#29141](#), Jason Dillaman)
- msg/async: enable secure mode by default, no longer experimental ([pr#29143](#), Sage Weil)
- msg/async: no-need set connection for Message ([pr#29142](#), Jianpeng Ma)
- msg/async, v2: make the reset\_recv\_state() unconditional ([issue#40115](#), [pr#29140](#), Radoslaw Zarzynski, Sage Weil)
- nautilus:common/options.cc: Lower the default value of osd\_deep\_scrub\_large\_omap\_object\_key\_threshold ([pr#29173](#), Neha Ojha)

- osd: Don't randomize deep scrubs when noscrub set ([issue#40198](#), [pr#28768](#), David Zafman)
- osd: Fix the way that auto repair triggers after regular scrub ([issue#40530](#), [issue#40073](#), [pr#28869](#), [sjust@redhat.com](#), David Zafman)
- osd/OSD: auto mark heartbeat sessions as stale and tear them down ([issue#40586](#), [pr#29391](#), xie xingguo)
- osd/OSD: keep synchronizing with mon if stuck at booting ([pr#28639](#), xie xingguo)
- osd/PG: do not queue scrub if PG is not active when unblock ([issue#40451](#), [pr#29372](#), Sage Weil)
- osd/PG: fix cleanup of pgmeta-like objects on PG deletion ([pr#29115](#), Sage Weil)
- pybind/mgr/rbd\_support: ignore missing support for RBD namespaces ([issue#41475](#), [pr#29945](#), Mykola Golub)
- rbd/action: fix error getting positional argument ([issue#40095](#), [pr#28870](#), songweibin)
- rbd: [cli] 'export' should handle concurrent IO completions ([issue#40435](#), [pr#29329](#), Jason Dillaman)
- rbd: librbd: do not unblock IO prior to growing object map during resize ([issue#39952](#), [pr#29246](#), Jason Dillaman)
- rbd-mirror: handle duplicates in image sync throttler queue ([issue#40519](#), [pr#28817](#), Mykola Golub)
- rbd-mirror: link against the specified alloc library ([issue#40110](#), [pr#29193](#), Jason Dillaman)
- rbd-nbd: sscnf return 0 mean not-match ([issue#39269](#), [pr#29315](#), Jianpeng Ma)
- rbd: profile rbd OSD cap should add class rbd metadata\_list cap by default ([issue#39973](#), [pr#29328](#), songweibin)
- rbd: Reduce log level for cls/journal and cls/rbd expected errors ([issue#40865](#), [pr#29551](#), Jason Dillaman)
- rbd: tests: add "rbd diff" coverage to suite ([issue#39447](#), [pr#28575](#), Shyukri Shyukriev, Nathan Cutler)
- rgw: add 'GET /admin/realm?list' api to list realms ([issue#39626](#), [pr#28751](#), Casey Bodley)
- rgw: allow radosgw-admin to list bucket w -allow-unordered ([issue#39637](#), [pr#28230](#), J. Eric Ivancich)
- rgw: conditionally allow builtin users with non-unique email addresses

([issue#40089](#), [pr#28715](#), Matt Benjamin)

- rgw: deleting bucket can fail when it contains unfinished multipart uploads ([issue#40526](#), [pr#29154](#), J. Eric Ivancich)
- rgw: Don't crash on copy when metadata directive not supplied ([issue#40416](#), [pr#29499](#), Adam C. Emerson)
- rgw\_file: advance\_mtime() should consider namespace expiration ([issue#40415](#), [pr#29410](#), Matt Benjamin)
- rgw\_file: advance\_mtime() takes RGWFileHandle::mutex unconditionally ([pr#29801](#), Matt Benjamin)
- rgw\_file: all directories are virtual with respect to contents ([issue#40204](#), [pr#28886](#), Matt Benjamin)
- rgw\_file: fix invalidation of top-level directories ([issue#40196](#), [pr#29309](#), Matt Benjamin)
- rgw\_file: fix readdir eof() calc-caller stop implies !eof ([issue#40375](#), [pr#29409](#), Matt Benjamin)
- rgw\_file: include tenant when hashing bucket names ([issue#40118](#), [pr#28854](#), Matt Benjamin)
- rgw: fix miss get ret in STSService::storeARN ([issue#40386](#), [pr#28713](#), Tianshan Qu)
- rgw: fix prefix handling in LCFilter ([issue#37879](#), [pr#28550](#), Matt Benjamin)
- rgw: fix rgw crash and set correct error code ([pr#28729](#), yuliyang)
- rgw: hadoop-s3a suite failing with more ansible errors ([issue#39706](#), [pr#28735](#), Casey Bodley)
- rgw: hadoop-s3a suite failing with more ansible errors ([issue#39706](#), [pr#29265](#), Casey Bodley)
- rgw: Librgw doesn't GC deleted object correctly ([issue#37734](#), [pr#28648](#), Tao Chen, Matt Benjamin)
- rgw: multisite: DELETE Bucket CORS is not forwarded to master zone ([issue#39629](#), [pr#28714](#), Chang Liu)
- rgw: multisite: fix -bypass-gc flag for 'radosgw-admin bucket rm' ([issue#24991](#), [pr#28549](#), Casey Bodley)
- rgw: multisite: 'radosgw-admin bilog trim' stops after 1000 entries ([issue#40187](#), [pr#29326](#), Casey Bodley)
- rgw: multisite: 'radosgw-admin bucket sync status' should call

- syncs\_from(source.name) instead of id ([issue#40022](#), [pr#28739](#), Casey Bodley)
- rgw: multisite: radosgw-admin commands should not modify metadata on a non-master zone ([issue#39548](#), [pr#29163](#), Shilpa Jagannath)
  - rgw: multisite: RGWListBucketIndexesCR for data full sync needs pagination ([issue#39551](#), [pr#29311](#), Shilpa Jagannath)
  - rgw/OutputDataSocket: append\_output(buffer::list&) says it will (but does not) discard output at data\_max\_backlog ([issue#40178](#), [pr#29310](#), Matt Benjamin)
  - rgw, Policy should be url\_decode when assume\_role ([pr#28728](#), yuliyang)
  - rgw: provide admin-friendly reshard status output ([issue#37615](#), [pr#29286](#), Mark Kogan)
  - rgw: Put LC doesn't clear existing lifecycle ([issue#39654](#), [pr#29313](#), Abhishek Lekshmanan)
  - rgw: remove rgw\_num\_rados\_handles; set autoscale parameters or rgw metadata pools ([pr#27684](#), Adam C. Emerson, Casey Bodley, Sage Weil)
  - rgw: RGWGC add perfcounter retire counter ([issue#38251](#), [pr#29308](#), Matt Benjamin)
  - rgw: Save an unnecessary copy of RGWEnv ([issue#40183](#), [pr#29205](#), Mark Kogan)
  - rgw: set null version object issues ([issue#36763](#), [pr#29287](#), Tianshan Qu)
  - rgw: Swift interface: server side copy fails if object name contains "?" ([issue#27217](#), [pr#28736](#), Casey Bodley)
  - rgw: TempURL should not allow PUTs with the X-Object-Manifest ([issue#20797](#), [pr#28712](#), Radoslaw Zarzynski)
  - rgw: the Multi-Object Delete operation of S3 API wrongly handles the Code response element ([issue#18241](#), [pr#28737](#), Radoslaw Zarzynski)
  - rocksdb: rocksdb\_rmrang related improvements ([pr#29439](#), Zengran Zhang, Sage Weil)
  - rocksdb: Updated to v6.1.2 ([pr#29440](#), Mark Nelson)
  - tools: ceph-kvstore-tool: print db stats ([pr#28810](#), Igor Fedotov)

## v14.2.2 Nautilus

---

This is the second bug fix release of Ceph Nautilus release series. We recommend all Nautilus users upgrade to this release. For upgrading from older releases of ceph, general guidelines for upgrade to nautilus must be followed [Upgrading from Mimic or Luminous](#).

# Notable Changes

- The no{up,down,in,out} related commands have been revamped. There are now 2 ways to set the no{up,down,in,out} flags: the old ‘ceph osd [un]set <flag>’ command, which sets cluster-wide flags; and the new ‘ceph osd [un]set-group <flags> <who>’ command, which sets flags in batch at the granularity of any crush node, or device class.
- radosgw-admin introduces two subcommands that allow the managing of expire-stale objects that might be left behind after a bucket reshards in earlier versions of RGW. One subcommand lists such objects and the other deletes them. Read the troubleshooting section of the dynamic resharding docs for details.
- Earlier Nautilus releases (14.2.1 and 14.2.0) have an issue where deploying a single new (Nautilus) BlueStore OSD on an upgraded cluster (i.e. one that was originally deployed pre-Nautilus) breaks the pool utilization stats reported by `ceph df`. Until all OSDs have been reprovisioned or updated (via `ceph-bluestore-tool repair`), the pool stats will show values that are lower than the true value. This is resolved in 14.2.2, such that the cluster only switches to using the more accurate per-pool stats after *all* OSDs are 14.2.2 (or later), are BlueStore, and (if they were created prior to Nautilus) have been updated via the `repair` function.
- The default value for `mon_crush_min_required_version` has been changed from firefly to hammer, which means the cluster will issue a health warning if your CRUSH tunables are older than hammer. There is generally a small (but non-zero) amount of data that will move around by making the switch to hammer tunables; for more information, see [Tunables](#).

If possible, we recommend that you set the oldest allowed client to hammer or later. You can tell what the current oldest allowed client is with:

```
1. ceph osd dump | grep min_compat_client
```

If the current value is older than hammer, you can tell whether it is safe to make this change by verifying that there are no clients older than hammer currently connected to the cluster with:

```
1. ceph features
```

The newer straw2 CRUSH bucket type was introduced in hammer, and ensuring that all clients are hammer or newer allows new features only supported for straw2 buckets to be used, including the crush-compat mode for the [Balancer](#).

# Changelog

---

- bluestore: backport more bluestore alerts ([pr#27645](#), Sage Weil, Igor Fedotov)
- bluestore: call fault\_range prior to looking for blob to reuse ([pr#27525](#), Igor Fedotov)
- bluestore: correctly measure deferred writes into new blobs ([issue#38816](#), [pr#27819](#), Sage Weil)
- bluestore: dump before “no-spanning blob id” abort ([pr#28028](#), Igor Fedotov)
- bluestore: fix for FreeBSD iocb structure ([issue#39612](#), [pr#28007](#), Willem Jan Withagen)
- bluestore: fix missing discard in BlueStore::\_kv\_sync\_thread ([issue#39672](#), [pr#28258](#), Junhui Tang)
- bluestore: fix out-of-bound access in bmap allocator ([pr#27740](#), Igor Fedotov)
- bluestore: fix duplicate allocations in bmap allocator ([issue#40080](#), [pr#28646](#), Igor Fedotov)
- build/ops: Ceph RPM build fails on openSUSE Tumbleweed with GCC 9 ([issue#40067](#), [issue#39974](#), [pr#28299](#), Martin Liška)
- build/ops: cmake: Fix build against ncurses with separate libtinfo ([pr#27532](#), Lars Wendler)
- build/ops: cmake: set empty-string RPATH for ceph-osd ([issue#40301](#), [issue#40295](#), [pr#28516](#), Nathan Cutler)
- build/ops: do\_cmake.sh: source not found ([issue#39981](#), [issue#40003](#), [pr#28215](#), Nathan Cutler)
- build/ops: python3 pybind RPMs do not replace their python2 counterparts on upgrade even though they should ([issue#40099](#), [issue#40232](#), [pr#28469](#), Nathan Cutler)
- build/ops: rpm: install grafana dashboards world readable ([pr#28392](#), Jan Fajerski)
- build/ops: selinux: Update the policy for RHEL8 ([pr#28511](#), Boris Ranto)
- ceph-volume: add utility functions ([pr#27791](#), Mohamad Gebai)
- ceph-volume: broken assertion errors after pytest changes ([pr#28925](#), Alfredo Deza)
- ceph-volume: look for rotational data in lsblk ([pr#27723](#), Andrew Schoen)

- ceph-volume: tests add a sleep in tox for slow OSDs after booting ([pr#28924](#), Alfredo Deza)
- ceph-volume: use the Device.rotational property instead of sys\_api ([pr#29028](#), Andrew Schoen)
- cephfs-shell: Revert “cephfs.pyx: add py3 compatibility” ([pr#28641](#), Varsha Rao)
- cephfs-shell: ls command produces error: no colorize attribute found error ([issue#39376](#), [issue#39378](#), [issue#38740](#), [issue#39379](#), [issue#39197](#), [issue#39377](#), [pr#27677](#), Milind Changire, Varsha Rao)
- cephfs-shell: misc. cephfs-shell backports ([issue#40314](#), [issue#40471](#), [issue#40418](#), [issue#40469](#), [issue#40313](#), [issue#39937](#), [issue#39678](#), [issue#40244](#), [issue#39404](#), [issue#40243](#), [issue#39165](#), [issue#40470](#), [issue#40455](#), [issue#39936](#), [issue#40217](#), [pr#28681](#), Patrick Donnelly, Varsha Rao, Milind Changire)
- cephfs-shell: mkdir error for relative path ([issue#39960](#), [pr#28616](#), Varsha Rao)
- cephfs: FSAL\_CEPH assertion failed in Client::\_lookup\_name: “parent->is\_dir()” ([issue#40085](#), [issue#40161](#), [pr#28612](#), Jeff Layton)
- cephfs: ceph\_volume\_client: Too many arguments for “WriteOpCtx” ([issue#39050](#), [issue#38946](#), [pr#27893](#), Ramana Raja)
- cephfs: client: ceph.dir.rctime xattr value incorrectly prefixes 09 to the nanoseconds component ([issue#40167](#), [pr#28500](#), David Disseldorp)
- cephfs: client: fix “ceph.snap.btime” vxattr value ([issue#40169](#), [pr#28499](#), David Disseldorp)
- cephfs: client: fix fuse client hang because its bad session PipeConnection ([issue#39686](#), [issue#39305](#), [pr#28375](#), Guan yunfei)
- cephfs: kclient: nofail option not supported ([issue#39232](#), [pr#27851](#), Kenneth Waegeman)
- cephfs: mds: Expose CephFS snapshot creation time to clients ([issue#39471](#), [pr#27901](#), David Disseldorp)
- cephfs: mds: MDSTableServer.cc: 83: FAILED assert(version == tid) ([issue#39211](#), [issue#38835](#), [pr#27853](#), “Yan, Zheng”)
- cephfs: mds: avoid sending too many osd requests at once after mds restarts ([issue#40028](#), [issue#40040](#), [pr#28582](#), simon gao)
- cephfs: mds: behind on trimming and “[dentry] was purgeable but no longer is!” ([issue#39222](#), [issue#38679](#), [pr#27879](#), “Yan, Zheng”)
- cephfs: mds: better output of ‘ceph health detail’ ([issue#39266](#), [pr#27846](#), Shen Hang’)

- cephfs: mds: check dir fragment to split dir if mkdir makes it oversized ([issue#39690](#), [pr#28394](#), Erqi Chen)
- cephfs: mds: check directory split after rename ([issue#39199](#), [issue#38994](#), [pr#27736](#), Shen Hang)
- cephfs: mds: drop reconnect message from non-existent session ([issue#39026](#), [issue#39192](#), [pr#27714](#), Shen Hang)
- cephfs: mds: fail to resolve snapshot name contains '\_' ([issue#39473](#), [pr#27849](#), "Yan, Zheng")
- cephfs: mds: fix 'is session in blacklist' check in Server::apply\_blacklist() ([issue#40236](#), [issue#40061](#), [pr#28618](#), "Yan, Zheng")
- cephfs: mds: fix corner case of replaying open sessions ([pr#28580](#), "Yan, Zheng")
- cephfs: mds: high debug logging with many subtrees is slow ([issue#38876](#), [pr#27892](#), Rishabh Dave)
- cephfs: mds: initialize cap\_revoke\_eviction\_timeout with conf ([issue#39209](#), [issue#38844](#), [pr#27842](#), simon gao)
- cephfs: mds: output lock state in format dump ([issue#39645](#), [issue#39670](#), [pr#28233](#), Zhi Zhang)
- cephfs: mds: reset heartbeat during long-running loops in recovery ([issue#40223](#), [pr#28611](#), "Yan, Zheng")
- cephfs: mds: there is an assertion when calling Beacon::shutdown() ([issue#39214](#), [issue#38822](#), [pr#27852](#), huanwen ren)
- cephfs: mount: key parsing fail when doing a remount ([issue#40164](#), [pr#28610](#), Luis Henriques)
- cephfs: pybind: added lseek() ([pr#28333](#), Xiaowei Chu)
- common/assert: include ceph\_abort\_msg(arg) arg in log output ([pr#27824](#), Sage Weil)
- common/options: annotate some options; enable some runtime updates ([pr#27818](#), Sage Weil)
- common/options: update mon\_crush\_min\_required\_version=hammer ([pr#27625](#), Sage Weil)
- common/util: handle long lines in /proc/cpuinfo ([issue#38296](#), [issue#39476](#), [pr#28141](#), Sage Weil)
- common: Clang requires a default constructor, but it can be empty ([issue#39561](#), [issue#39573](#), [pr#28131](#), Willem Jan Withagen)

- common: fix parse\_env nullptr deref ([pr#28382](#), Patrick Donnelly)
- common: make cluster\_network work ([issue#39671](#), [pr#28248](#), Jianpeng Ma)
- common: parse ISO 8601 datetime format ([issue#40087](#), [pr#28325](#), Sage Weil)
- core: Give recovery for inactive PGs a higher priority ([issue#39504](#), [issue#38195](#), [pr#27854](#), David Zafman)
- core: mon,osd: add no{out,down,in,out} flags on CRUSH nodes ([pr#27623](#), xie xingguo, Sage Weil)
- core: mon/Elector: format mon\_release correctly ([issue#39419](#), [pr#27771](#), Sage Weil)
- core: mon/Monitor: allow probe if MMonProbe::mon\_release == 0 ([issue#38850](#), [pr#28262](#), Sage Weil)
- core: mon: fix off-by-one rendering progress bar ([pr#28398](#), Sage Weil)
- core: mon: use per-pool stats only when all OSDs are reporting ([pr#29032](#), Sage Weil)
- core: monitoring: Provide a base set of Prometheus alert manager rules that notify the user about common Ceph error conditions ([issue#39540](#), [pr#27998](#), Jan Fajerski)
- core: monitoring: update Grafana dashboards ([issue#39652](#), [issue#40006](#), [issue#39971](#), [issue#39932](#), [pr#28101](#), Kiefer Chang, Jan Fajerski)
- core: osd/OSD.cc: make osd bench description consistent with parameters ([issue#39006](#), [issue#39375](#), [pr#28035](#), Neha Ojha)
- core: osd/OSDMap: Replace get\_out\_osds with get\_out\_existing\_osds ([issue#39421](#), [issue#39154](#), [pr#28072](#), Brad Hubbard)
- core: osd/PG: discover missing objects when an OSD peers and PG is degraded ([pr#27744](#), Jonas Jelten)
- core: osd/PG: do not use approx\_missing\_objects pre-nautilus ([issue#39512](#), [pr#28160](#), Neha Ojha)
- core: osd/PG: fix last\_complete re-calculation on splitting ([issue#39539](#), [issue#26958](#), [pr#28219](#), xie xingguo)
- core: osd/PG: skip rollforward when !transaction\_applied during append\_log() ([issue#36739](#), [issue#38881](#), [pr#27654](#), Neha Ojha)
- core: osd/PGLog: preserve original\_crt to check rollbackability ([issue#36739](#), [issue#39043](#), [pr#27632](#), Neha Ojha)
- core: osd: Don't evict after a flush if intersecting scrub range ([issue#38840](#),

- issue#39519, pr#28205, David Zafman')
- core: osd: Don't include user changeable flag in snaptrim related assert (issue#39699, issue#38124, pr#28203, David Zafman')
- core: osd: FAILED ceph\_assert(attrs || !pg\_log.get\_missing().is\_missing(soid) || (it\_objects != pg\_log.get\_log().objects.end() && it\_objects->second->op == pg\_log\_entry\_t::LOST\_REVERT)) in PrimaryLogPG::get\_object\_context() (issue#38931, issue#39219, issue#38784, pr#27839, xie xingguo)
- core: osd: Include dups in copy\_after() and copy\_up\_to() (issue#39304, pr#28088, David Zafman)
- core: osd: Increase log level of messages which unnecessarily fill up logs (pr#27687, David Zafman)
- core: osd: Output Base64 encoding of CRC header if binary data present (issue#39738, pr#28504, David Zafman)
- core: osd: Primary won't automatically repair replica on pulling error (issue#39101, issue#39184, pr#27711, xie xingguo, David Zafman')
- core: osd: revamp {noup, nodown, noin, noout} related commands (pr#28400, xie xingguo)
- core: osd: shutdown recovery\_request\_timer earlier (issue#39205, pr#27803, Zengran Zhang)
- core: osd: take heartbeat\_lock when calling heartbeat() (issue#39514, issue#39439, pr#28164, Sage Weil)
- doc: add LAZYIO (issue#39051, issue#38729, pr#27899, "Yan, Zheng")
- doc: add documentation for "fs set min\_compat\_client" (issue#39130, issue#39176, pr#27900, Patrick Donnelly)
- doc: cleanup HTTP Frontends documentation (issue#38874, pr#27922, Casey Bodley)
- doc: dashboard documentation changes (pr#27642, Tatjana Dehler, Lenz Grimmer)
- doc: orchestrator\_cli: Rook orch supports mon update (issue#39169, issue#39137, pr#27488, Sebastian Wagner)
- doc: osd\_internals/async\_recovery: update cost calculation (pr#28046, Neha Ojha)
- doc: rados/operations/devices: document device prediction (pr#27752, Sage Weil)
- mgr/ActivePyModules: handle\_command - fix broken lock (issue#39235, issue#39308, pr#27939, xie xingguo)
- mgr/BaseMgrModule: run MonCommandCompletion on the finisher (issue#39397, issue#39335, pr#27699, Sage Weil)

- mgr/ansible: Host ls implementation ([issue#39559](#), [pr#27919](#), Juan Miguel Olmo Martxc3xadnez)
- mgr/balancer: various compat weight-set fixes ([pr#28279](#), xie xingguo)
- mgr/dashboard: Add custom dialogue for configuring PG scrub parameters ([issue#40059](#), [pr#28555](#), Tatjana Dehler)
- mgr/dashboard: Admin resource not honored ([issue#39338](#), [issue#39467](#), [pr#27868](#), Wido den Hollander)
- mgr/dashboard: Angular is creating multiple instances of the same service ([issue#39996](#), [issue#40075](#), [pr#28312](#), Tiago Melo)
- mgr/dashboard: Avoid merge conflicts in messages.xlf by auto-generating it at build time? ([issue#39658](#), [pr#28178](#), Sebastian Krah)
- mgr/dashboard: Display correct dialog title ([pr#28189](#), Volker Theile)
- mgr/dashboard: Error creating NFS client without squash ([issue#40074](#), [pr#28311](#), Tiago Melo)
- mgr/dashboard: KV-table transforms dates through pipe ([issue#39558](#), [pr#28021](#), Stephan Mxc3xbcller)
- mgr/dashboard: Localization for date picker module ([issue#39371](#), [pr#27673](#), Stephan Mxc3xbcller)
- mgr/dashboard: Manager should complain about wrong dashboard certificate ([issue#39346](#), [pr#27742](#), Volker Theile)
- mgr/dashboard: NFS clients information is not displayed in the details view ([issue#40057](#), [pr#28318](#), Tiago Melo)
- mgr/dashboard: NFS export creation: Add more info to the validation message of the field Pseudo ([issue#39975](#), [issue#39327](#), [pr#28320](#), Tiago Melo)
- mgr/dashboard: Only one root node is shown in the crush map viewer ([issue#39647](#), [issue#40077](#), [pr#28316](#), Tiago Melo)
- mgr/dashboard: Push Grafana dashboards on startup ([pr#28635](#), Zack Cerza)
- mgr/dashboard: Queue notifications as default ([issue#39560](#), [pr#28022](#), Stephan Mxc3xbcller)
- mgr/dashboard: RBD snapshot name suggestion with local time suffix ([issue#39534](#), [pr#27890](#), Stephan Mxc3xbcller)
- mgr/dashboard: Reduce the number of renders on the tables ([issue#39944](#), [issue#40076](#), [pr#28315](#), Tiago Melo)
- mgr/dashboard: Some validations are not updated and prevent the submission of a

- ```
form (issue#40030, pr#28319, Tiago Melo)
```
- mgr/dashboard: Unable to see tcmu-runner perf counters ([issue#39988](#), [pr#28191](#), Ricardo Marques)
  - mgr/dashboard: Unify the look of dashboard charts ([issue#39384](#), [issue#39961](#), [pr#28175](#), Tiago Melo)
  - mgr/dashboard: Validate if any client belongs to more than one group ([issue#39036](#), [issue#39454](#), [pr#27760](#), Tiago Melo)
  - mgr/dashboard: code documentation ([issue#39345](#), [issue#36243](#), [pr#27746](#), Ernesto Puerta)
  - mgr/dashboard: iSCSI GET requests should not be logged ([pr#28024](#), Ricardo Marques)
  - mgr/dashboard: iSCSI form does not support IPv6 ([pr#28026](#), Ricardo Marques)
  - mgr/dashboard: iSCSI form is showing a warning ([issue#39452](#), [issue#39324](#), [pr#27758](#), Tiago Melo)
  - mgr/dashboard: iSCSI should allow exporting an RBD image with Journaling enabled ([pr#28011](#), Ricardo Marques)
  - mgr/dashboard: inconsistent result when editing a RBD image's features ([issue#39993](#), [issue#39933](#), [pr#28218](#), Kiefer Chang')
  - mgr/dashboard: incorrect help message for minimum blob size ([issue#39624](#), [issue#39664](#), [pr#28062](#), Kiefer Chang)
  - mgr/dashboard: local variable 'cluster\_id' referenced before assignment error when trying to list NFS Ganesha daemons ([issue#40031](#), [pr#28261](#), Nur Faizin')
  - mgr/dashboard: make auth token work with UTC times only ([issue#39524](#), [issue#39300](#), [pr#27942](#), Ricardo Dias)
  - mgr/dashboard: openssl exception when verifying certificates of HTTPS requests ([issue#39962](#), [issue#39628](#), [pr#28163](#), Ricardo Dias)
  - mgr/dashboard: orchestrator mgr modules assert failure on iscsi service request ([issue#40037](#), [pr#28552](#), Sebastian Wagner)
  - mgr/dashboard: show degraded/misplaced/unfound objects ([pr#28584](#), Alfonso Martxc3xadnez)
  - mgr/orchestrator: Remove "(add|test|remove)\_stateful\_service\_rule" ([issue#38808](#), [pr#27043](#), Sebastian Wagner)
  - mgr/orchestrator: add progress events to all orchestrators ([pr#28040](#), Sebastian Wagner)

- mgr/progress: behave if pgs disappear (due to a racing pg merge) ([issue#38157](#), [issue#39344](#), [pr#27608](#), Sage Weil)
- mgr/prometheus: replace whitespaces in metrics' names ([pr#27886](#), Alfonso Martxc3xadnez')
- mgr/rook: Added missing rgw daemons in service ls ([issue#39171](#), [issue#39312](#), [pr#27864](#), Sebastian Wagner)
- mgr/rook: Fix RGW creation ([issue#39158](#), [issue#39313](#), [pr#27863](#), Sebastian Wagner)
- mgr/rook: Remove support for Rook older than v0.9 ([issue#39356](#), [issue#39278](#), [pr#27862](#), Sebastian Wagner)
- mgr/test\_orchestrator: AttributeError: 'TestWriteCompletion' object has no attribute 'id' ([issue#39536](#), [pr#27920](#), Sebastian Wagner')
- mgr/volumes: FS subvolumes enhancements ([issue#40429](#), [pr#28767](#), Ramana Raja)
- mgr/volumes: add CephFS subvolumes library ([issue#39750](#), [issue#40152](#), [issue#39949](#), [issue#40014](#), [issue#39610](#), [pr#28429](#), Sage Weil, Venky Shankar, Ramana Raja, Rishabh Dave)
- mgr/volumes: refactor volume module ([issue#40378](#), [issue#39969](#), [pr#28595](#), Venky Shankar)
- mgr: Update the restful module in nautilus ([pr#28291](#), Kefu Chai, Boris Ranto)
- mgr: deadlock ([issue#39040](#), [issue#39425](#), [pr#28098](#), xie xingguo)
- mgr: fix pgp\_num adjustments ([issue#38626](#), [pr#27876](#), Sage Weil, Marius Schiffer)
- mgr: log an error if we can't find any modules to load ([issue#40090](#), [pr#28347](#), Tim Serong')
- monitoring: pybind/mgr: fix format for rbd-mirror prometheus metrics ([pr#28485](#), Mykola Golub)
- msg/async: connection race + winner fault can leave connection stuck at replacing forever ([issue#39241](#), [issue#37499](#), [issue#39448](#), [issue#38493](#), [pr#27915](#), Jason Dillaman, xie xingguo)
- msg/async/ProtocolV[12]: add ms\_learn\_addr\_from\_peer ([pr#28589](#), Sage Weil)
- msg: output peer address when detecting bad CRCs ([issue#39367](#), [pr#27857](#), Greg Farnum)
- pybind: Add 'RBD\_FEATURE\_MIGRATING' to rbd.pyx ([issue#39609](#), [issue#39736](#), [pr#28482](#), Ricardo Marques')
- pybind: Rados.get\_fsid() returning bytes in python3 ([issue#40192](#), [issue#38381](#), [pr#28476](#), Jason Dillaman)

- rbd: krbd: fix rbd map hang due to udev return subsystem unordered ([issue#39089](#), [issue#39315](#), [pr#28019](#), Zhi Zhang)
- rbd: librbd: async open/close should free ImageCtx before issuing callback ([issue#39428](#), [issue#39031](#), [pr#28121](#), Jason Dillaman)
- rbd: librbd: avoid dereferencing an empty container during deep-copy ([issue#40368](#), [issue#40379](#), [pr#28577](#), Jason Dillaman)
- rbd: librbd: do not allow to deep copy migrating image ([issue#39224](#), [pr#27882](#), Mykola Golub)
- rbd: librbd: fix issues with object-map/fast-diff feature interlock ([issue#39946](#), [issue#39521](#), [pr#28127](#), Jason Dillaman)
- rbd: librbd: fixed several race conditions related to copyup ([issue#39195](#), [issue#39021](#), [pr#28132](#), Jason Dillaman)
- rbd: librbd: make flush be queued by QOS throttler ([issue#38869](#), [pr#28120](#), Mykola Golub)
- rbd: librbd: re-add support for nautilus clients talking to jewel clusters ([issue#39450](#), [pr#27936](#), Jason Dillaman)
- rbd: librbd: support EC data pool images sparsify ([issue#39226](#), [pr#27903](#), Mykola Golub)
- rbd: rbd-mirror: clear out bufferlist prior to listing mirror images ([issue#39462](#), [issue#39407](#), [pr#28122](#), Jason Dillaman)
- rbd: rbd-mirror: image replayer should periodically flush IO and commit positions ([issue#39257](#), [issue#39288](#), [pr#27937](#), Jason Dillaman)
- rgw: Evaluating bucket policies also while reading permissions for anxe2x80xa6 ([issue#38638](#), [issue#39273](#), [pr#27918](#), Pritha Srivastava)
- rgw: admin: handle delete\_at attr in object stat output ([pr#27827](#), Abhishek Lekshmanan)
- rgw: beast: multiple v4 and v6 endpoints with the same port will cause failure ([issue#39746](#), [issue#39038](#), [pr#28541](#), Abhishek Lekshmanan)
- rgw: beast: set a default port for endpoints ([issue#39048](#), [issue#39000](#), [pr#27660](#), Abhishek Lekshmanan)
- rgw: bucket stats report mtime in UTC ([pr#27826](#), Alfonso Martxc3xadnez, Casey Bodley)
- rgw: clean up some logging ([issue#39503](#), [pr#27953](#), J. Eric Ivancich)
- rgw: cloud sync module fails to sync multipart objects ([issue#39684](#), [pr#28064](#),

Abhishek Lekshmanan)

- rgw: cloud sync module logs attrs in the log ([issue#39574](#), [pr#27954](#), Nathan Cutler)
- rgw: crypto: throw DigestException from Digest and HMAC ([issue#39676](#), [issue#39456](#), [pr#28309](#), Matt Benjamin)
- rgw: document CreateBucketConfiguration for s3 PUT Bucket request ([issue#39597](#), [issue#39601](#), [pr#28512](#), Casey Bodley)
- rgw: fix Multisite sync corruption ([pr#28383](#), Tianshan Qu, Casey Bodley, Xiaoxi CHEN)
- rgw: fix bucket may redundantly list keys after BI\_PREFIX\_CHAR ([issue#39984](#), [issue#40148](#), [pr#28410](#), Casey Bodley, Tianshan Qu)
- rgw: fix default\_placement containing "/" when storage\_class is standard ([issue#39745](#), [issue#39380](#), [pr#28538](#), mkogan1)
- rgw: inefficient unordered bucket listing ([issue#39410](#), [issue#39393](#), [pr#27924](#), Casey Bodley)
- rgw: librgw: unexpected crash when creating bucket ([issue#39575](#), [pr#27955](#), Tao CHEN)
- rgw: limit entries in remove\_olh\_pending\_entries() ([issue#39178](#), [issue#39118](#), [pr#27664](#), Casey Bodley)
- rgw: list bucket with start marker and delimiter will miss next object with char '0' ([issue#40762](#), [issue#39989](#), [pr#29022](#), Tianshan Qu)
- rgw: multisite log trimming only checks peers that sync from us ([issue#39283](#), [pr#27814](#), Casey Bodley)
- rgw: multisite: add perf counters to data sync ([issue#38549](#), [issue#38918](#), [pr#27921](#), Abhishek Lekshmanan, Casey Bodley)
- rgw: multisite: mismatch of bucket creation times from List Buckets ([issue#39635](#), [issue#39735](#), [pr#28444](#), Casey Bodley)
- rgw: multisite: period pusher gets 403 Forbidden against other zonegroups ([issue#39287](#), [issue#39414](#), [pr#27952](#), Casey Bodley)
- rgw: race condition between resharding and ops waiting on resharding ([issue#39202](#), [pr#27800](#), J. Eric Ivancich)
- rgw: radosgw-admin: add tenant argument to reshards cancel ([issue#39018](#), [pr#27630](#), Abhishek Lekshmanan)
- rgw: rgw\_file: save etag and acl info in setattr ([issue#39228](#), [pr#27904](#), Tao

Chen)

- rgw: swift object expiry fails when a bucket reshards ([issue#39740](#), [pr#28537](#), Abhishek Lekshmanan)
- rgw: unittest\_rgw\_dmclock\_scheduler does not need Boost\_LIBRARIES ([issue#39577](#), [pr#27944](#), Willem Jan Withagen)
- rgw: update resharding documentation ([issue#39046](#), [pr#27923](#), J. Eric Ivancich)
- tests: added bluestore\_warn\_on\_legacy\_statfs: false setting ([issue#40467](#), [pr#28723](#), Yuri Weinstein)
- tests: added ragweed coverage to stress-split\\* upgrade suites ([issue#40452](#), [issue#40467](#), [pr#28661](#), Yuri Weinstein)
- tests: added v14.2.1 ([issue#40181](#), [pr#28416](#), Yuri Weinstein)
- tests: cannot schedule kcephfs/multimds ([issue#40116](#), [pr#28369](#), Patrick Donnelly)
- tests: centos 7.6 etc ([pr#27439](#), Sage Weil)
- tests: ceph-ansible: ceph-ansible requires ansible 2.8 ([issue#40602](#), [issue#40669](#), [pr#28871](#), Brad Hubbard)
- tests: ceph-ansible: cephfs\_pools variable pgs should be pg\_num ([issue#40670](#), [issue#40605](#), [pr#28872](#), Brad Hubbard)
- tests: cephfs-shell: teuthology tests ([issue#39935](#), [issue#39526](#), [pr#28614](#), Milind Changire)
- tests: cephfs: TestMisc.test\_evict\_client fails ([issue#40220](#), [pr#28613](#), "Yan, Zheng")
- tests: cleaned up supported distro for nautilus ([pr#28065](#), Yuri Weinstein)
- tests: ignore legacy bluestore stats errors ([issue#40374](#), [pr#28563](#), Patrick Donnelly)
- tests: librbd: drop 'ceph\_test\_librbd\_api' target ([issue#39423](#), [issue#39072](#), [pr#28091](#), Jason Dillaman')
- tests: mgr: tox failures when running make check ([issue#39323](#), [issue#39530](#), [pr#27884](#), Nathan Cutler)
- tests: pass -ssh-config to pytest to resolve hosts when connecting ([pr#28923](#), Alfredo Deza)
- tests: rbd: qemu-io tests fail under latest Ubuntu kernel ([issue#39541](#), [issue#24668](#), [pr#27988](#), Jason Dillaman)
- tests: removed 1node and systemd tests as ceph-deploy is not axe2x80xa6

([pr#28458](#), Yuri Weinstein)

- tests: rgw: fix race in test\_rgwreshard\_wait and test\_rgwrushard\_wait uses same clock for timing ([issue#39479](#), [pr#27779](#), Casey Bodley)
- tests: rgw: fix swift warning message ([issue#40304](#), [pr#28698](#), Casey Bodley)
- tests: rgw: more fixes for swift task ([issue#40304](#), [pr#28922](#), Casey Bodley)
- tests: rgw: skip swift tests on rhel 7.6+ ([issue#40402](#), [issue#40304](#), [pr#28604](#), Casey Bodley)
- tests: stop testing simple messenger in fs qa ([issue#40373](#), [pr#28562](#), Patrick Donnelly)
- tests: tasks/rbd\_fio: fixed missing delimiter between 'cd' and 'configure' ([issue#39590](#), [pr#27989](#), Jason Dillaman)
- tests: test\_sessionmap assumes simple messenger ([issue#39430](#), [pr#27772](#), Patrick Donnelly)
- tests: use curl in wait\_for\_radosgw() in util/rgw.py ([issue#40346](#), [pr#28598](#), Ali Maredia)
- tests: workunits/rbd: use https protocol for devstack git operations ([issue#39656](#), [issue#39729](#), [pr#28128](#), Jason Dillaman)
- tests: workunits/rbd: wait for rbd-nbd unmap to complete ([issue#39675](#), [issue#39598](#), [pr#28273](#), Jason Dillaman)

## v14.2.1 Nautilus

---

This is the first bug fix release of Ceph Nautilus release series. We recommend all nautilus users upgrade to this release. For upgrading from older releases of ceph, general guidelines for upgrade to nautilus must be followed [Upgrading from Mimic or Luminous](#).

## Notable Changes

---

- Ceph now packages python bindings for python3.6 instead of python3.4, because EPEL7 recently switched from python3.4 to python3.6 as the native python3. see the [announcement](#) for more details on the background of this change.

## Known Issues

---

- Nautilus-based librbd clients cannot open images stored on pre-Luminous clusters

# Changelog

- bluestore: ceph-bluestore-tool: bluefs-bdev-expand cmd might assert if no WAL is configured ([issue#39253](#), [pr#27523](#), Igor Fedotov)
- bluestore: os/bluestore: fix bitmap allocator issues ([pr#27139](#), Igor Fedotov)
- build/ops,rgw: rgw: build async scheduler only when beast is built ([pr#27191](#), Abhishek Lekshmanan)
- build/ops: build/ops: Running ceph under Pacemaker control not supported by SUSE Linux Enterprise ([issue#38862](#), [pr#27127](#), Nathan Cutler)
- build/ops: build/ops: ceph-mgr-diskprediction-local requires numpy and scipy on SUSE, but these packages do not exist on SUSE ([issue#38863](#), [pr#27125](#), Nathan Cutler)
- build/ops: cmake/FindRocksDB: fix IMPORTED\_LOCATION for ROCKSDB\_LIBRARIES ([issue#38993](#), [pr#27601](#), dudengke)
- build/ops: cmake: revert librados\_tp.so version from 3 to 2 ([issue#39291](#), [issue#39293](#), [pr#27597](#), Nathan Cutler)
- build/ops: qa,rpm,cmake: switch over to python3.6 ([issue#39236](#), [issue#39164](#), [pr#27505](#), Boris Ranto, Kefu Chai)
- cephfs: fs: we lack a feature bit for nautilus ([issue#39078](#), [issue#39187](#), [pr#27497](#), Patrick Donnelly)
- cephfs: ls -S command produces AttributeError: 'str' object has no attribute 'decode' ([pr#27531](#), Varsha Rao)
- cephfs: mds|kclient: MDS\_CLIENT\_LATE\_RELEASE warning caused by inline bug on RHEL 7.5 ([issue#39225](#), [pr#27500](#), "Yan, Zheng")
- common,core: crush: various fixes for weight-sets, the osd\_crush\_update\_weight\_set option, and tests ([pr#27119](#), Sage Weil)
- common/blkdev: get\_device\_id: behave if model is lvm and id\_model\_enc isn't there ([pr#27158](#), Sage Weil)
- common/config: parse -default-\$option as a default value ([pr#27217](#), Sage Weil)
- core,mgr: mgr: autoscale down can lead to max\_pg\_per\_osd limit ([issue#39271](#), [issue#38786](#), [pr#27547](#), Sage Weil)
- core,mon: mon/Monitor.cc: print min\_mon\_release correctly ([pr#27168](#), Neha Ojha)
- core,tests: tests: osd-markdown.sh can fail with CLI\_DUP\_COMMAND=1 ([issue#38359](#), [issue#39275](#), [pr#27550](#), Sage Weil)

- core: Improvements to auto repair ([issue#38616](#), [pr#27220](#), xie xingguo, David Zafman)
- core: Rook: Fix creation of Bluestore OSDs ([issue#39167](#), [issue#39062](#), [pr#27486](#), Sebastian Wagner)
- core: ceph-objectstore-tool: rename dump-import to dump-export ([issue#39325](#), [issue#39284](#), [pr#27610](#), David Zafman)
- core: common/blkdev: handle devices with ID\_MODEL as "LVM PV ..." but valid ID\_MODEL\_ENC ([pr#27096](#), Sage Weil)
- core: common: fix deferred log starting ([pr#27388](#), Sage Weil, Jason Dillaman)
- core: crush/CrushCompiler: Fix \_\_replacement\_assert ([issue#39174](#), [pr#27620](#), Brad Hubbard)
- core: global: explicitly call out EIO events in crash dumps ([pr#27440](#), Sage Weil)
- core: log: log\_to\_file + -default-\* + fixes and improvements ([pr#27278](#), Sage Weil)
- core: mon/MgrStatMonitor: ensure only one copy of initial service map ([issue#38839](#), [pr#27116](#), Sage Weil)
- core: mon/OSDMonitor: allow 'osd pool set pgp\_num\_actual' ([pr#27060](#), Sage Weil)
- core: mon: make mon\_osd\_down\_out\_subtree\_limit update at runtime ([pr#27582](#), Sage Weil)
- core: mon: ok-to-stop commands for mon and mds ([pr#27347](#), Sage Weil)
- core: mon: quiet devname log noise ([pr#27314](#), Sage Weil)
- core: osd/OSDMap: add 'zone' to default crush map ([pr#27117](#), Sage Weil)
- core: osd/PGLog.h: print olog\_can\_rollback\_to before deciding to rollback ([issue#38906](#), [issue#38894](#), [pr#27302](#), Neha Ojha)
- core: osd/osd\_types: fix object\_stat\_sum\_t fast-path decode ([issue#39320](#), [issue#39281](#), [pr#27555](#), David Zafman)
- core: osd: backport recent upmap fixes ([issue#38860](#), [issue#38967](#), [issue#38897](#), [issue#38826](#), [pr#27225](#), huangjun, xie xingguo)
- core: osd: process\_copy\_chunk remove obc ref before pg unlock ([issue#38842](#), [issue#38973](#), [pr#27478](#), Zengran Zhang)
- dashboard: NFS: failed to disable NFSv3 in export create ([issue#39104](#), [issue#38997](#), [pr#27368](#), Tiago Melo)
- doc/releases/nautilus: fix config update step ([pr#27502](#), Sage Weil)

- doc: doc/orchestrator: Fix broken bullet points ([issue#39168](#), [pr#27487](#), Sebastian Wagner)
- doc: doc: Minor rados related documentation fixes ([issue#38896](#), [issue#38903](#), [pr#27189](#), David Zafman)
- doc: doc: rgw: Added library/package for Golang ([issue#38730](#), [issue#38867](#), [pr#27549](#), Irek Fasikhov)
- install-deps.sh: install ‘\*rpm-macros’ ([issue#39164](#), [pr#27544](#), Kefu Chai)
- mgr/dashboard add polish language ([issue#39052](#), [pr#27287](#), Sebastian Krah)
- mgr/dashboard/qa: Improve tasks.mgr.test\_dashboard.TestDashboard.test\_standby ([pr#27237](#), Volker Theile)
- mgr/dashboard: 1 osds exist in the crush map but not in the osdmap breaks OSD page ([issue#38885](#), [issue#36086](#), [pr#27543](#), Patrick Nawracay)
- mgr/dashboard: Adapt iSCSI overview page to make use of ceph-iscsi ([pr#27541](#), Ricardo Marques)
- mgr/dashboard: Add date range and log search functionality ([issue#37387](#), [issue#38878](#), [pr#27283](#), guodan1)
- mgr/dashboard: Add refresh interval to the dashboard landing page ([issue#26872](#), [issue#38988](#), [pr#27267](#), guodan1)
- mgr/dashboard: Add separate option to config SSL port ([issue#39001](#), [pr#27393](#), Volker Theile)
- mgr/dashboard: Added breadcrumb tests to NFS menu ([issue#38981](#), [pr#27589](#), Nathan Weinberg)
- mgr/dashboard: Back button component ([issue#39058](#), [pr#27405](#), Stephan Müller)
- mgr/dashboard: Cannot submit NFS export form when NFSv4 is not selected ([issue#39105](#), [issue#39063](#), [pr#27370](#), Tiago Melo)
- mgr/dashboard: Error creating NFS export without UDP ([issue#39107](#), [issue#39090](#), [pr#27372](#), Tiago Melo)
- mgr/dashboard: Error on iSCSI disk diff ([pr#27460](#), Ricardo Marques)
- mgr/dashboard: Filter iSCSI target images based on required features ([issue#39002](#), [pr#27363](#), Ricardo Marques)
- mgr/dashboard: Fix env vars of run-tox.sh ([issue#38798](#), [issue#38864](#), [pr#27361](#), Patrick Nawracay)
- mgr/dashboard: Fixes tooltip behavior ([pr#27395](#), Stephan Müller)

- mgr/dashboard: FixtureHelper ([issue#39041](#), [pr#27398](#), Stephan Müller)
- mgr/dashboard: NFS Squash field should be required ([issue#39106](#), [issue#39064](#), [pr#27371](#), Tiago Melo)
- mgr/dashboard: PreventDefault isn't working on 400 errors ([pr#27389](#), Stephan Müller)
- mgr/dashboard: Typo in "CephFS Name" field on NFS form ([issue#39067](#), [pr#27449](#), Tiago Melo)
- mgr/dashboard: dashboard giving 401 unauthorized ([issue#38871](#), [pr#27219](#), ming416)
- mgr/dashboard: fix sparkline component ([issue#38866](#), [pr#27260](#), Alfonso Martínez)
- mgr/dashboard: readonly user can't see any pages ([issue#39240](#), [pr#27611](#), Stephan Müller)
- mgr/dashboard: unify button/URL actions naming + bugfix (add whitelist to guard) ([issue#37337](#), [issue#39003](#), [pr#27492](#), Ernesto Puerta)
- mgr/dashboard: update vstart to use new ssl\_server\_port ([issue#39124](#), [pr#27394](#), Ernesto Puerta)
- mgr/deepsea: use ceph\_volume output in get\_inventory() ([issue#39083](#), [pr#27319](#), Tim Serong)
- mgr/diskprediction\_cloud: Correct base64 encode translate table ([pr#27167](#), Rick Chen)
- mgr/orchestrator: Add error handling to interface ([issue#38837](#), [pr#27095](#), Sebastian Wagner)
- mgr/pg\_autoscaler: add pg\_autoscale\_bias ([pr#27387](#), Sage Weil)
- mgr: mgr/dashboard: Error on iSCSI target submission ([pr#27461](#), Ricardo Marques)
- mgr: ceph-mgr: ImportError: Interpreter change detected - this module can only be loaded into one interpreter per process ([issue#38865](#), [pr#27128](#), Tim Serong)
- mgr: mgr/DaemonServer: handle\_conf\_change - fix broken locking ([issue#38964](#), [issue#38899](#), [pr#27454](#), xie xingguo)
- mgr: mgr/balancer: Python 3 compatibility fix ([issue#38831](#), [issue#38855](#), [pr#27227](#), Marius Schiffer)
- mgr: mgr/dashboard: Check if gateway is in use before allowing the deletion via iscsi-gateway-rm command ([pr#27457](#), Ricardo Marques)
- mgr: mgr/dashboard: Display the number of active sessions for each iSCSI target ([pr#27450](#), Ricardo Marques)

- mgr: mgr/devicehealth: Fix python 3 incompatibility ([issue#38957](#), [issue#38939](#), [pr#27390](#), Marius Schiffer)
- mgr: mgr/telemetry: add report\_timestamp to sent reports ([pr#27701](#), Dan Mick)
- mgr: mgr/telemetry: use list; redact host; 24h default interval ([pr#27709](#), Sage Weil, Dan Mick)
- mgr: mgr: Configure Py root logger for Mgr modules ([issue#38969](#), [pr#27261](#), Volker Theile)
- mgr: mgr: Diskprediction unable to transfer data into the cloud server ([issue#38970](#), [pr#27240](#), Rick Chen)
- mon/MonClient: do not dereference auth\_supported.end() ([pr#27215](#), Kefu Chai)
- mon/MonmapMonitor: clean up empty created stamp in monmap ([issue#39085](#), [pr#27399](#), Sage Weil)
- mon: mon: add cluster log to file option ([pr#27346](#), Sage Weil)
- msg/async v2: make v2 work on rdma ([pr#27216](#), Jianpeng Ma)
- msg: default to debug\_ms=0 ([pr#27197](#), Sage Weil)
- osd: OSDMapRef access by multiple threads is unsafe ([pr#27402](#), Zengran Zhang, Kefu Chai)
- qa/valgrind ([pr#27320](#), Radoslaw Zarzynski)
- rbd,tests: backport krbd discard qa fixes to nautilus ([issue#38861](#), [pr#27258](#), Ilya Dryomov)
- rbd,tests: backport krbd discard qa fixes to stable branches ([issue#38956](#), [pr#27239](#), Ilya Dryomov)
- rbd: librbd: ignore -EOPNOTSUPP errors when retrieving image group membership ([issue#38834](#), [pr#27080](#), Jason Dillaman)
- rbd: librbd: look for pool metadata in default namespace ([issue#38961](#), [pr#27423](#), Mykola Golub)
- rbd: librbd: trash move return EBUSY instead of EINVAL for migrating image ([issue#38968](#), [pr#27475](#), Mykola Golub)
- rbd: rbd: krbd: return -ETIMEDOUT in polling ([issue#38792](#), [issue#38977](#), [pr#27539](#), Dongsheng Yang)
- rgw: Adding tcp\_nodelay option to Beast ([issue#38926](#), [pr#27355](#), Or Friedmann)
- rgw: Fix S3 compatibility bug when CORS is not found ([issue#38923](#), [issue#37945](#), [pr#27331](#), Nick Janus)

- rgw: LC: handle resharded buckets ([pr#27559](#), Abhishek Lekshmanan)
- rgw: Make rgw admin ops api get user info consistent with the command line ([issue#39135](#), [pr#27501](#), Li Shuhao)
- rgw: don't crash on missing /etc/mime.types ([issue#38921](#), [issue#38328](#), [pr#27329](#), Casey Bodley)
- rgw: don't recalculate etags for slo/dlo ([pr#27561](#), Casey Bodley)
- rgw: fix RGWDeleteMultiObj::verify\_permission() ([issue#38980](#), [pr#27586](#), Irek Fasikhov)
- rgw: fix read not exists null version return wrong ([issue#38811](#), [issue#38909](#), [pr#27306](#), Tianshan Qu)
- rgw: ldap: fix early return in LDAPAuthEngine::init w/uri not empty() ([issue#38754](#), [pr#26972](#), Matt Benjamin)
- rgw: multisite: data sync loops back to the start of the datalog after reaching the end ([issue#39075](#), [issue#39033](#), [pr#27498](#), Casey Bodley)
- rgw: nfs: skip empty (non-POSIX) path segments ([issue#38744](#), [issue#38773](#), [pr#27208](#), Matt Benjamin)
- rgw: nfs: svc-enable RGWLib ([issue#38774](#), [pr#27232](#), Matt Benjamin)
- rgw: orphans find perf improvements ([issue#39181](#), [pr#27560](#), Abhishek Lekshmanan)
- rgw: rgw admin: disable stale instance deletion in multisite ([issue#39015](#), [pr#27602](#), Abhishek Lekshmanan)
- rgw: sse c fixes ([issue#38700](#), [pr#27296](#), Adam Kupczyk, Casey Bodley, Abhishek Lekshmanan)
- rgw: support delimiter longer than one symbol ([issue#38777](#), [pr#27548](#), Matt Benjamin)
- rook-ceph-system namespace hardcoded in the rook orchestrator ([issue#38799](#), [issue#39250](#), [pr#27496](#), Sebastian Wagner)
- rpm,cmake: use specified python3 version if any ([pr#27382](#), Kefu Chai)

## v14.2.0 Nautilus

---

This is the first stable release of Ceph Nautilus.

## Major Changes from Mimic

---

- *Dashboard:*

The [Ceph Dashboard](#) has gained a lot of new functionality:

- Support for multiple users / roles
- SSO (SAMLv2) for user authentication
- Auditing support
- New landing page, showing more metrics and health info
- I18N support
- REST API documentation with Swagger API

New Ceph management features include:

- OSD management (mark as down/out, change OSD settings, recovery profiles)
- Cluster config settings editor
- Ceph Pool management (create/modify/delete)
- ECP management
- RBD mirroring configuration
- Embedded Grafana Dashboards (derived from Ceph Metrics)
- CRUSH map viewer
- NFS Ganesha management
- iSCSI target management (via [Ceph iSCSI Gateway](#))
- RBD QoS configuration
- Ceph Manager (ceph-mgr) module management
- Prometheus alert Management

Also, the Ceph Dashboard is now split into its own package named [ceph-mgr-dashboard](#) . You might want to install it separately, if your package management software fails to do so when it installs [ceph-mgr](#) .

- *RADOS:*

- The number of placement groups (PGs) per pool can now be decreased at any time, and the cluster can [automatically tune the PG count](#) based on cluster utilization or administrator hints.
- The new [v2 wire protocol](#) brings support for encryption on the wire.

- Physical [storage devices](#) consumed by OSD and Monitor daemons are now tracked by the cluster along with health metrics (i.e., SMART), and the cluster can apply a pre-trained prediction model or a cloud-based prediction service to [warn about expected HDD or SSD failures](#).
- The NUMA node for OSD daemons can easily be monitored via the `ceph osd numa-status` command, and configured via the `osd_numa_node` config option.
- When BlueStore OSDs are used, space utilization is now broken down by object data, omap data, and internal metadata, by pool, and by pre- and post-compression sizes.
- OSDs more effectively prioritize the most important PGs and objects when performing recovery and backfill.
- Progress for long-running background processes-like recovery after a device failure-is now reported as part of `ceph status` .
- An experimental [Coupled-Layer “Clay” erasure code](#) plugin has been added that reduces network bandwidth and IO needed for most recovery operations.
- *RGW:*
  - S3 lifecycle transition for tiering between storage classes.
  - A new web frontend (Beast) has replaced civetweb as the default, improving overall performance.
  - A new publish/subscribe infrastructure allows RGW to feed events to serverless frameworks like knative or data pipelines like Kafka.
  - A range of authentication features, including STS federation using OAuth2 and OpenID::connect and an OPA (Open Policy Agent) authentication delegation prototype.
  - The new archive zone federation feature enables full preservation of all objects (including history) in a separate zone.
- *CephFS:*
  - MDS stability has been greatly improved for large caches and long-running clients with a lot of RAM. Cache trimming and client capability recall is now throttled to prevent overloading the MDS.
  - CephFS may now be exported via NFS-Ganesha clusters in environments managed by Rook. Ceph manages the clusters and ensures high-availability and scalability. An [introductory demo](#) is available. More automation of this feature is expected to be forthcoming in future minor releases of Nautilus.
  - The MDS `mds_standby_for_*` , `mon_force_standby_active` , and `mds_standby_replay` configuration options have been obsoleted. Instead, the operator [may now set](#)

the new `allow_standby_replay` flag on the CephFS file system. This setting causes standbys to become standby-replay for any available rank in the file system.

- MDS now supports dropping its cache which concurrently asks clients to trim their caches. This is done using MDS admin socket `cache drop` command.
- It is now possible to check the progress of an on-going scrub in the MDS. Additionally, a scrub may be paused or aborted. See [the scrub documentation](#) for more information.
- A new interface for creating volumes is provided via the `ceph volume` command-line-interface.
- A new `cephfs-shell` tool is available for manipulating a CephFS file system without mounting.
- CephFS-related output from `ceph status` has been reformatted for brevity, clarity, and usefulness.
- Lazy IO has been revamped. It can be turned on by the client using the new `CEPH_O_LAZY` flag to the `ceph_open` C/C++ API or via the config option `client_force_lazyio`.
- CephFS file system can now be brought down rapidly via the `ceph fs fail` command. See [the administration page](#) for more information.
- *RBD:*
  - Images can be live-migrated with minimal downtime to assist with moving images between pools or to new layouts.
  - New `rbd perf image iotop` and `rbd perf image iostat` commands provide an iotop- and iostat-like IO monitor for all RBD images.
  - The `ceph-mgr` Prometheus exporter now optionally includes an IO monitor for all RBD images.
  - Support for separate image namespaces within a pool for tenant isolation.
- *Misc:*
  - Ceph has a new set of [orchestrator modules](#) to directly interact with external orchestrators like ceph-ansible, DeepSea, Rook, or simply ssh via a consistent CLI (and, eventually, Dashboard) interface.

# Upgrading from Mimic or Luminous

## Notes

- During the upgrade from Luminous to Nautilus, it will not be possible to create a new OSD using a Luminous ceph-osd daemon after the monitors have been upgraded to Nautilus. We recommend you avoid adding or replacing any OSDs while the upgrade is in progress.
- We recommend you avoid creating any RADOS pools while the upgrade is in progress.
- You can monitor the progress of your upgrade at each stage with the `ceph versions` command, which will tell you what ceph version(s) are running for each type of daemon.

## Instructions

- If your cluster was originally installed with a version prior to Luminous, ensure that it has completed at least one full scrub of all PGs while running Luminous. Failure to do so will cause your monitor daemons to refuse to join the quorum on start, leaving them non-functional.

If you are unsure whether or not your Luminous cluster has completed a full scrub of all PGs, you can check your cluster's state by running:

```
1. # ceph osd dump | grep ^flags
```

In order to be able to proceed to Nautilus, your OSD map must include the `recovery_deletes` and `purged_snapdirs` flags.

If your OSD map does not contain both these flags, you can simply wait for approximately 24-48 hours, which in a standard cluster configuration should be ample time for all your placement groups to be scrubbed at least once, and then repeat the above process to recheck.

However, if you have just completed an upgrade to Luminous and want to proceed to Mimic in short order, you can force a scrub on all placement groups with a one-line shell command, like:

```
1. # ceph pg dump pgs_brief | cut -d " " -f 1 | xargs -n1 ceph pg scrub
```

You should take into consideration that this forced scrub may possibly have a negative impact on your Ceph clients' performance.

- Make sure your cluster is stable and healthy (no down or recovering OSDs).

(Optional, but recommended.)

3. Set the `noout` flag for the duration of the upgrade. (Optional, but recommended.):

```
1. # ceph osd set noout
```

4. Upgrade monitors by installing the new packages and restarting the monitor daemons. For example, on each monitor host,:

```
1. # systemctl restart ceph-mon.target
```

Once all monitors are up, verify that the monitor upgrade is complete by looking for the `nautilus` string in the mon map. The command:

```
1. # ceph mon dump | grep min_mon_release
```

should report:

```
1. min_mon_release 14 (nautilus)
```

If it doesn't, that implies that one or more monitors hasn't been upgraded and restarted and/or the quorum does not include all monitors.

5. Upgrade `ceph-mgr` daemons by installing the new packages and restarting all manager daemons. For example, on each manager host,:

```
1. # systemctl restart ceph-mgr.target
```

Please note, if you are using Ceph Dashboard, you will probably need to install `ceph-mgr-dashboard` separately after upgrading `ceph-mgr` package. The install script of `ceph-mgr-dashboard` will restart the manager daemons automatically for you. So in this case, you can just skip the step to restart the daemons.

Verify the `ceph-mgr` daemons are running by checking `ceph -s` :

```
1. # ceph -s
2.
3. ...
4. services:
5.   mon: 3 daemons, quorum foo,bar,baz
6.   mgr: foo(active), standbys: bar, baz
7. ...
```

6. Upgrade all OSDs by installing the new packages and restarting the `ceph-osd` daemons on all OSD hosts:

```
1. # systemctl restart ceph-osd.target
```

You can monitor the progress of the OSD upgrades with the `ceph versions` or `ceph osd versions` commands:

```
1. # ceph osd versions
2. {
3.     "ceph version 13.2.5 (...) mimic (stable)": 12,
4.     "ceph version 14.2.0 (...) nautilus (stable)": 22,
5. }
```

7. If there are any OSDs in the cluster deployed with ceph-disk (e.g., almost any OSDs that were created before the Mimic release), you need to tell ceph-volume to adopt responsibility for starting the daemons. On each host containing OSDs, ensure the OSDs are currently running, and then:

```
1. # ceph-volume simple scan
2. # ceph-volume simple activate --all
```

We recommend that each OSD host be rebooted following this step to verify that the OSDs start up automatically.

Note that ceph-volume doesn't have the same hot-plug capability that ceph-disk did, where a newly attached disk is automatically detected via udev events. If the OSD isn't currently running when the above `scan` command is run, or a ceph-disk-based OSD is moved to a new host, or the host OSD is reinstalled, or the `/etc/ceph/osd` directory is lost, you will need to scan the main data partition for each ceph-disk OSD explicitly. For example,:

```
1. # ceph-volume simple scan /dev/sdb1
```

The output will include the appropriate `ceph-volume simple activate` command to enable the OSD.

8. Upgrade all CephFS MDS daemons. For each CephFS file system,

- i. Reduce the number of ranks to 1. (Make note of the original number of MDS daemons first if you plan to restore it later.):

```
1. # ceph status
2. # ceph fs set <fs_name> max_mds 1
```

- ii. Wait for the cluster to deactivate any non-zero ranks by periodically checking the status:

```
1. # ceph status
```

iii. Take all standby MDS daemons offline on the appropriate hosts with:

```
1. # systemctl stop ceph-mds@<daemon_name>
```

iv. Confirm that only one MDS is online and is rank 0 for your FS:

```
1. # ceph status
```

v. Upgrade the last remaining MDS daemon by installing the new packages and restarting the daemon:

```
1. # systemctl restart ceph-mds.target
```

vi. Restart all standby MDS daemons that were taken offline:

```
1. # systemctl start ceph-mds.target
```

vii. Restore the original value of `max_mds` for the volume:

```
1. # ceph fs set <fs_name> max_mds <original_max_mds>
```

9. Upgrade all radosgw daemons by upgrading packages and restarting daemons on all hosts:

```
1. # systemctl restart ceph-radosgw.target
```

10. Complete the upgrade by disallowing pre-Nautilus OSDs and enabling all new Nautilus-only functionality:

```
1. # ceph osd require-osd-release nautilus
```

### Important

This step is mandatory. Failure to execute this step will make it impossible for OSDs to communicate after msgrv2 is enabled.

11. If you set `noout` at the beginning, be sure to clear it with:

```
1. # ceph osd unset noout
```

12. Verify the cluster is healthy with `ceph health`.

If your CRUSH tunables are older than Hammer, Ceph will now issue a health warning. If you see a health alert to that effect, you can revert this change with:

```
1. ceph config set mon mon_crush_min_required_version firefly
```

If Ceph does not complain, however, then we recommend you also switch any existing CRUSH buckets to straw2, which was added back in the Hammer release. If you have any ‘straw’ buckets, this will result in a modest amount of data movement, but generally nothing too severe.:

```
1. ceph osd getcrushmap -o backup-crushmap
2. ceph osd crush set-all-straw-buckets-to-straw2
```

If there are problems, you can easily revert with:

```
1. ceph osd setcrushmap -i backup-crushmap
```

Moving to ‘straw2’ buckets will unlock a few recent features, like the crush-compat [balancer](#) mode added back in Luminous.

13. To enable the new [v2 network protocol](#), issue the following command:

```
1. ceph mon enable-msgr2
```

This will instruct all monitors that bind to the old default port 6789 for the legacy v1 protocol to also bind to the new 3300 v2 protocol port. To see if all monitors have been updated,:

```
1. ceph mon dump
```

and verify that each monitor has both a `v2:` and `v1:` address listed.

Running nautilus OSDs will not bind to their v2 address automatically. They must be restarted for that to happen.

#### Important

Before this step is run, the following command must already have been run:

```
# ceph osd require-osd-release nautilus
```

If this command (step 10 in this procedure) has not been run, OSDs will lose the ability to communicate.

14. For each host that has been upgraded, you should update your `ceph.conf` file so that it either specifies no monitor port (if you are running the monitors on the default ports) or references both the v2 and v1 addresses and ports explicitly. Things will still work if only the v1 IP and port are listed, but each CLI instantiation or daemon will need to reconnect after learning the monitors also speak the v2 protocol, slowing things down a bit and preventing a full transition

to the v2 protocol.

This is also a good time to fully transition any config options in `ceph.conf` into the cluster's configuration database. On each host, you can use the following command to import any options into the monitors with:

```
1. ceph config assimilate-conf -i /etc/ceph/ceph.conf
```

You can see the cluster's configuration database with:

```
1. ceph config dump
```

To create a minimal but sufficient `ceph.conf` for each host,:

```
1. ceph config generate-minimal-conf > /etc/ceph/ceph.conf.new  
2. mv /etc/ceph/ceph.conf.new /etc/ceph/ceph.conf
```

Be sure to use this new config only on hosts that have been upgraded to Nautilus, as it may contain a `mon_host` value that includes the new `v2:` and `v1:` prefixes for IP addresses that is only understood by Nautilus.

For more information, see [Updating ceph.conf and mon\\_host](#).

15. Consider enabling the `telemetry` module to send anonymized usage statistics and crash information to the Ceph upstream developers. To see what would be reported (without actually sending any information to anyone),:

```
1. ceph mgr module enable telemetry  
2. ceph telemetry show
```

If you are comfortable with the data that is reported, you can opt-in to automatically report the high-level cluster metadata with:

```
1. ceph telemetry on
```

For more information about the `telemetry` module, see [the documentation](#).

# Upgrading from pre-Luminous releases (like Jewel)

You *must* first upgrade to Luminous (12.2.z) before attempting an upgrade to Nautilus. In addition, your cluster must have completed at least one scrub of all PGs while running Luminous, setting the `recovery_deletes` and `purged_snapdirs` flags in the OSD map.

## Upgrade compatibility notes

These changes occurred between the Mimic and Nautilus releases.

- `ceph pg stat` output has been modified in json format to match `ceph df` output:
  - “raw\_bytes” field renamed to “total\_bytes”
  - “raw\_bytes\_avail” field renamed to “total\_bytes\_avail”
  - “raw\_bytes\_avail” field renamed to “total\_bytes\_avail”
  - “raw\_bytes\_used” field renamed to “total\_bytes\_raw\_used”
  - “total\_bytes\_used” field added to represent the space (accumulated over all OSDs) allocated purely for data objects kept at block(slow) device
- `ceph df [detail]` output (GLOBAL section) has been modified in plain format:
  - new ‘USED’ column shows the space (accumulated over all OSDs) allocated purely for data objects kept at block(slow) device.
  - ‘RAW USED’ is now a sum of ‘USED’ space and space allocated/reserved at block device for Ceph purposes, e.g. BlueFS part for BlueStore.
- `ceph df [detail]` output (GLOBAL section) has been modified in json format:
  - ‘total\_used\_bytes’ column now shows the space (accumulated over all OSDs) allocated purely for data objects kept at block(slow) device
  - new ‘total\_used\_raw\_bytes’ column shows a sum of ‘USED’ space and space allocated/reserved at block device for Ceph purposes, e.g. BlueFS part for BlueStore.
- `ceph df [detail]` output (POOLS section) has been modified in plain format:
  - ‘BYTES USED’ column renamed to ‘STORED’. Represents amount of data stored by the user.
  - ‘USED’ column now represent amount of space allocated purely for data by all

- OSD nodes in KB.
- ‘QUOTA BYTES’, ‘QUOTA OBJECTS’ aren’t showed anymore in non-detailed mode.
- new column ‘USED COMPR’ - amount of space allocated for compressed data. i.e., compressed data plus all the allocation, replication and erasure coding overhead.
- new column ‘UNDER COMPR’ - amount of data passed through compression (summed over all replicas) and beneficial enough to be stored in a compressed form.
- Some columns reordering
- `ceph df [detail]` output (POOLS section) has been modified in json format:
  - ‘bytes used’ column renamed to ‘stored’. Represents amount of data stored by the user.
  - ‘raw bytes used’ column renamed to “stored\_raw”. Totals of user data over all OSD excluding degraded.
  - new ‘bytes\_used’ column now represent amount of space allocated by all OSD nodes.
  - ‘kb\_used’ column - the same as ‘bytes\_used’ but in KB.
  - new column ‘compress\_bytes\_used’ - amount of space allocated for compressed data. i.e., compressed data plus all the allocation, replication and erasure coding overhead.
  - new column ‘compress\_under\_bytes’ amount of data passed through compression (summed over all replicas) and beneficial enough to be stored in a compressed form.
- `rados df [detail]` output (POOLS section) has been modified in plain format:
  - ‘USED’ column now shows the space (accumulated over all OSDs) allocated purely for data objects kept at block(slow) device.
  - new column ‘USED COMPR’ - amount of space allocated for compressed data. i.e., compressed data plus all the allocation, replication and erasure coding overhead.
  - new column ‘UNDER COMPR’ - amount of data passed through compression (summed over all replicas) and beneficial enough to be stored in a compressed form.
- `rados df [detail]` output (POOLS section) has been modified in json format:
  - ‘size\_bytes’ and ‘size\_kb’ columns now show the space (accumulated over all OSDs) allocated purely for data objects kept at block device.

- new column ‘compress\_bytes\_used’ - amount of space allocated for compressed data. i.e., compressed data plus all the allocation, replication and erasure coding overhead.
- new column ‘compress\_under\_bytes’ amount of data passed through compression (summed over all replicas) and beneficial enough to be stored in a compressed form.
- `ceph pg dump` output (totals section) has been modified in json format:
  - new ‘USED’ column shows the space (accumulated over all OSDs) allocated purely for data objects kept at block(slow) device.
  - ‘USED\_RAW’ is now a sum of ‘USED’ space and space allocated/reserved at block device for Ceph purposes, e.g. BlueFS part for BlueStore.
- The `ceph osd rm` command has been deprecated. Users should use `ceph osd destroy` or `ceph osd purge` (but after first confirming it is safe to do so via the `ceph osd safe-to-destroy` command).
- The MDS now supports dropping its cache for the purposes of benchmarking.:

```
1. ceph tell mds.* cache drop <timeout>
```

Note that the MDS cache is cooperatively managed by the clients. It is necessary for clients to give up capabilities in order for the MDS to fully drop its cache. This is accomplished by asking all clients to trim as many caps as possible. The timeout argument to the `cache drop` command controls how long the MDS waits for clients to complete trimming caps. This is optional and is 0 by default (no timeout). Keep in mind that clients may still retain caps to open files which will prevent the metadata for those files from being dropped by both the client and the MDS. (This is an equivalent scenario to dropping the Linux page/buffer/inode/dentry caches with some processes pinning some inodes/dentries/pages in cache.)

- The `mon_health_preluminous_compat` and `mon_health_preluminous_compat_warning` config options are removed, as the related functionality is more than two versions old. Any legacy monitoring system expecting Jewel-style health output will need to be updated to work with Nautilus.
- Nautilus is not supported on any distros still running upstart so upstart specific files and references have been removed.
- The `ceph pg <pgid> list_missing` command has been renamed to `ceph pg <pgid> list_unfound` to better match its behaviour.
- The *rbd-mirror* daemon can now retrieve remote peer cluster configuration secrets from the monitor. To use this feature, the rbd-mirror daemon CephX user for the local cluster must use the `profile rbd-mirror` mon cap. The secrets can be set using

the `rbd mirror pool peer add` and `rbd mirror pool peer set` actions.

- The ‘rbd-mirror’ daemon will now run in active/active mode by default, where mirrored images are evenly distributed between all active ‘rbd-mirror’ daemons. To revert to active/pассив mode, override the ‘`rbd_mirror_image_policy_type`’ config key to ‘none’.
- The `ceph mds deactivate` is fully obsolete and references to it in the docs have been removed or clarified.
- The libcephfs bindings added the `ceph_select_filesystem` function for use with multiple filesystems.
- The cephfs python bindings now include `mount_root` and `filesystem_name` options in the `mount()` function.
- erasure-code: add experimental *Coupled LAYER (CLAY)* erasure codes support. It features less network traffic and disk I/O when performing recovery.
- The `cache drop` OSD command has been added to drop an OSD’s caches:
  - `ceph tell osd.x cache drop`
- The `cache status` OSD command has been added to get the cache stats of an OSD:
  - `ceph tell osd.x cache status`
- The libcephfs added several functions that allow restarted client to destroy or reclaim state held by a previous incarnation. These functions are for NFS servers.
- The `ceph` command line tool now accepts keyword arguments in the format `--arg=value` or `--arg value`.
- `librados::IoCtx::nobjects_begin()` and `librados::NObjectIterator` now communicate errors by throwing a `std::system_error` exception instead of `std::runtime_error`.
- The callback function passed to `LibRGWFS.readdir()` now accepts a `flags` parameter. it will be the last parameter passed to `readdir()` method.
- The `cephfs-data-scan scan_links` now automatically repair inotables and snaptable.
- Configuration values `mon_warn_not_scrubbed` and `mon_warn_not_deep_scrubbed` have been renamed. They are now `mon_warn_pg_not_scrubbed_ratio` and `mon_warn_pg_not_deep_scrubbed_ratio` respectively. This is to clarify that these warnings are related to pg scrubbing and are a ratio of the related interval. These options are now enabled by default.
- The MDS cache trimming is now throttled. Dropping the MDS cache via the `ceph tell mds.<foo> cache drop` command or large reductions in the cache size will no longer

cause service unavailability.

- The CephFS MDS behavior with recalling caps has been significantly improved to not attempt recalling too many caps at once, leading to instability. MDS with a large cache (64GB+) should be more stable.
- MDS now provides a config option `mds_max_caps_per_client` (default: 1M) to limit the number of caps a client session may hold. Long running client sessions with a large number of caps have been a source of instability in the MDS when all of these caps need to be processed during certain session events. It is recommended to not unnecessarily increase this value.
- The MDS config `mds_recall_state_timeout` has been removed. Late client recall warnings are now generated based on the number of caps the MDS has recalled which have not been released. The new configs `mds_recall_warning_threshold` (default: 32K) and `mds_recall_warning_decay_rate` (default: 60s) sets the threshold for this warning.
- The Telegraf module for the Manager allows for sending statistics to an Telegraf Agent over TCP, UDP or a UNIX Socket. Telegraf can then send the statistics to databases like InfluxDB, ElasticSearch, Graphite and many more.
- The graylog fields naming the originator of a log event have changed: the string-form name is now included (e.g., `"name": "mgr.foo"`), and the rank-form name is now in a nested section (e.g., `"rank": {"type": "mgr", "num": 43243}`).
- If the cluster log is directed at syslog, the entries are now prefixed by both the string-form name and the rank-form name (e.g., `mgr.x mgr.12345 ...` instead of just `mgr.12345 ...`).
- The JSON output of the `ceph osd find` command has replaced the `ip` field with an `addrs` section to reflect that OSDs may bind to multiple addresses.
- CephFS clients without the 's' flag in their authentication capability string will no longer be able to create/delete snapshots. To allow `client.foo` to create/delete snapshots in the `bar` directory of filesystem `cephfs_a`, use command:
  - `ceph auth caps client.foo mon 'allow r' osd 'allow rw tag cephfs data=cephfs_a' mds 'allow rw, allow rws path=/bar'`
- The `osd_heartbeat_addr` option has been removed as it served no (good) purpose: the OSD should always check heartbeats on both the public and cluster networks.
- The `rados` tool's `mkpool` and `rmpool` commands have been removed because they are redundant; please use the `ceph osd pool create` and `ceph osd pool rm` commands instead.
- The `auid` property for cephx users and RADOS pools has been removed. This was an undocumented and partially implemented capability that allowed cephx users to map capabilities to RADOS pools that they "owned". Because there are no users we have

removed this support. If any cephx capabilities exist in the cluster that restrict based on auid then they will no longer parse, and the cluster will report a health warning like:

```
1. AUTH_BAD_CAPS 1 auth entities have invalid capabilities
2. client.bad osd capability parse failed, stopped at 'allow rwx auid 123' of 'allow rwx auid 123'
```

The capability can be adjusted with the `ceph auth caps` command. For example,:

```
1. ceph auth caps client.bad osd 'allow rwx pool foo'
```

- The `ceph-kvstore-tool repair` command has been renamed `destructive-repair` since we have discovered it can corrupt an otherwise healthy rocksdb database. It should be used only as a last-ditch attempt to recover data from an otherwise corrupted store.
- The default memory utilization for the mons has been increased somewhat. Rocksdb now uses 512 MB of RAM by default, which should be sufficient for small to medium-sized clusters; large clusters should tune this up. Also, the `mon_osd_cache_size` has been increase from 10 OSDMaps to 500, which will translate to an additional 500 MB to 1 GB of RAM for large clusters, and much less for small clusters.
- The `mgr/balancer/max_misplaced` option has been replaced by a new global `target_max_misplaced_ratio` option that throttles both balancer activity and automated adjustments to `pgp_num` (normally as a result of `pg_num` changes). If you have customized the balancer module option, you will need to adjust your config to set the new global option or revert to the default of .05 (5%).
- By default, Ceph no longer issues a health warning when there are misplaced objects (objects that are fully replicated but not stored on the intended OSDs). You can reenable the old warning by setting `mon_warn_on_misplaced` to `true`.
- The `ceph-create-keys` tool is now obsolete. The monitors automatically create these keys on their own. For now the script prints a warning message and exits, but it will be removed in the next release. Note that `ceph-create-keys` would also write the admin and bootstrap keys to /etc/ceph and /var/lib/ceph, but this script no longer does that. Any deployment tools that relied on this behavior should instead make use of the `ceph auth export <entity-name>` command for whichever key(s) they need.
- The `mon_osd_pool_ec_fast_read` option has been renamed `osd_pool_default_ec_fast_read` to be more consistent with other `osd_pool_default_*` options that affect default values for newly created RADOS pools.
- The `mon addr` configuration option is now deprecated. It can still be used to specify an address for each monitor in the `ceph.conf` file, but it only affects cluster creation and bootstrapping, and it does not support listing multiple

addresses (e.g., both a v2 and v1 protocol address). We strongly recommend the option be removed and instead a single `mon host` option be specified in the `[global]` section to allow daemons and clients to discover the monitors.

- New command `ceph fs fail` has been added to quickly bring down a file system. This is a single command that unsets the joinable flag on the file system and brings down all of its ranks.
- The `cache drop` admin socket command has been removed. The `ceph tell mds.X cache drop` remains.

# Detailed Changelog

- add monitoring subdir and Grafana cluster dashboard ([pr#21850](#), Jan Fajerski)
- auth,common: include cleanups ([pr#23774](#), Kefu Chai)
- bluestore: bluestore/NVMEDevice.cc: fix ceph\_assert() when enable SPDK with 64KB kernel page size ([issue#36624](#), [pr#24817](#), tone.zhang)
- bluestore: bluestore/NVMEDevice.cc: fix NVMEManager thread hang ([issue#37720](#), [pr#25646](#), tone.zhang, Steve Capper)
- bluestore: bluestore/NVMe: use PCIe selector as the path name ([pr#24144](#), Kefu Chai)
- bluestore,cephfs,core,rbd,rgw: buffer,denc: use ptr::const\_iterator for decode ([pr#22015](#), Kefu Chai, Casey Bodley)
- bluestore: ceph-kvstore-tool: dump fixes ([pr#25262](#), Adam Kupczyk)
- bluestore: common/blkdev: check retval of stat() ([pr#26040](#), Kefu Chai)
- bluestore,core: ceph-dencoder: add bluefs types ([pr#22463](#), Sage Weil)
- bluestore,core,mon,performance: osd,mon: enable level\_compaction\_dynamic\_level\_bytes for rocksdb ([issue#24361](#), [pr#22337](#), Kefu Chai)
- bluestore,core: os/bluestore: don't store/use path\_block.{db,wal} from meta ([pr#22462](#), Sage Weil, Alfredo Deza)
- bluestore: os/bluestore: add bluestore\_ignore\_data\_csum option ([pr#26233](#), Sage Weil)
- bluestore: os/bluestore: add boundary check for cache-autotune related settings ([issue#37507](#), [pr#25421](#), xie xingguo)
- bluestore: os/bluestore/BlueFS: only flush dirty devices when do \_fsync ([pr#22110](#), Jianpeng Ma)
- bluestore: os/bluestore: bluestore\_buffer\_hit\_bytes perf counter doesn't reset ([pr#23576](#), Igor Fedotov)
- bluestore: os/bluestore: check return value of \_open\_bluefs ([pr#25471](#), Jianpeng Ma)
- bluestore: os/bluestore: cleanups ([pr#22556](#), Jianpeng Ma)
- bluestore: os/bluestore: deep fsck fails on inspecting very large onodes ([pr#26170](#), Igor Fedotov)

- bluestore: os/bluestore: do not assert on non-zero err codes from compress() call ([pr#25891](#), Igor Fedotov)
- bluestore: os/bluestore: firstly delete db then delete bluefs if open db met error ([pr#22336](#), Jianpeng Ma)
- bluestore: os/bluestore: fix and unify log output on allocation failure ([pr#25335](#), Igor Fedotov)
- bluestore: os/bluestore: fix assertion in StupidAllocator::get\_fragmentation ([pr#23606](#), Igor Fedotov)
- bluestore: os/bluestore: fix bloom filter num entry miscalculation in repairer ([issue#25001](#), [pr#24076](#), Igor Fedotov)
- bluestore: os/bluestore: fix bluefs extent miscalculations on small slow device ([pr#22563](#), Igor Fedotov)
- bluestore: os/bluestore: fix race between remove\_collection and object removals ([pr#23257](#), Igor Fedotov)
- bluestore: os/bluestore: fixup access a destroy cond cause deadlock or undefined behavior ([pr#25659](#), linbing)
- bluestore: os/bluestore: introduce new BlueFS perf counter to track the amount of ([pr#22086](#), Igor Fedotov)
- bluestore: os/bluestore/KernelDevice: misc cleanup ([pr#21491](#), Jianpeng Ma)
- bluestore: os/bluestore/KernelDevice: use flock(2) for block device lock ([issue#38150](#), [pr#26245](#), Sage Weil)
- bluestore: os/bluestore: misc cleanup ([pr#22472](#), Jianpeng Ma)
- bluestore: os/bluestore: Only use F\_SET\_FILE\_RW\_HINT when available ([pr#26431](#), Willem Jan Withagen)
- bluestore: os/bluestore: Only use `WRITE_LIFE` when available ([pr#25735](#), Willem Jan Withagen)
- bluestore: os/bluestore: remove redundant fault\_range ([pr#22898](#), Jianpeng Ma)
- bluestore: os/bluestore: remove useless condition ([pr#22335](#), Jianpeng Ma)
- bluestore: os/bluestore: simplify and fix SharedBlob::put() ([issue#24211](#), [pr#22123](#), Sage Weil)
- bluestore: os/bluestore: support for FreeBSD ([pr#25608](#), Alan Somers, Kefu Chai)
- bluestore: osd/osd\_types: fix pg\_t::contains() to check pool id too ([issue#32731](#), [pr#24085](#), Sage Weil)

- bluestore: os/objectstore: add a new op OP\_CREATE ([pr#22385](#), Jianpeng Ma)
- bluestore,performance: common/PriorityCache: First Step toward priority based caching ([pr#22009](#), Mark Nelson)
- bluestore,performance: os/bluestore: allocator pruning ([pr#21854](#), Igor Fedotov)
- bluestore,performance: os/bluestore/BlueFS: reduce bufferlist rebuilds during WAL writes ([pr#21689](#), Piotr Dałek)
- bluestore,performance: os/bluestore: use the monotonic clock for perf counters latencies ([pr#22121](#), Mohamad Gebai)
- bluestore: silence Clang warning on possible uninitialized usage ([pr#25702](#), Willem Jan Withagen)
- bluestore: spdk: fix ceph-osd crash when activate SPDK ([issue#24371](#), [pr#22356](#), tone-zhang)
- bluestore: test/fio: add option single\_pool\_mode in ceph-bluestore.fio ([pr#21929](#), Jianpeng Ma)
- bluestore,tests: test/objectstore: fix random generator in allocator\_bench ([pr#22544](#), Igor Fedotov)
- bluestore,tools: os/bluestore: allow ceph-bluestore-tool to coalesce, add and migrate BlueFS backing volumes ([pr#23103](#), Igor Fedotov)
- bluestore,tools: tools/ceph-bluestore-tool: avoid mon/config access when calling global... ([pr#22085](#), Igor Fedotov)
- build/ops: Add new OpenSUSE Leap id for install-deps.sh ([issue#25064](#), [pr#22793](#), Kyr Shatskyy)
- build/ops: arch/arm: Allow ceph\_crc32c\_aarch64 to be chosen only if it is compil... ([pr#24126](#), David Wang)
- build/ops: auth: do not use GSS/KRB5 if ! HAVE\_GSSAPI ([pr#25460](#), Kefu Chai)
- build/ops: build: 32 bit architecture fixes ([pr#23485](#), James Page)
- build/ops: build: further removal of subman configuration ([issue#38261](#), [pr#26368](#), Alfredo Deza)
- build/ops: build: LLVM ld does not like the versioning scheme ([pr#26801](#), Willem Jan Withagen)
- build/ops: ceph-create-keys: Misc Python 3 fixes ([issue#37641](#), [pr#25411](#), James Page)
- build/ops,cephfs: deb,rpm: fix python-cephfs dependencies ([issue#24919](#), [issue#24918](#), [pr#23043](#), Kefu Chai)

- build/ops: ceph.in: Add support for python 3 ([pr#24739](#), Tiago Melo)
- build/ops: ceph.spec.in: Don't use noarch for mgr module subpackages, fix /usr/lib64/ceph/mgr dir ownership ([pr#26398](#), Tim Serong)
- build/ops: change ceph-mgr package depency from py-bcrypt to python2-bcrypt ([issue#27206](#), [pr#23648](#), Konstantin Sakhinov)
- build/ops: civetweb: pull up to ceph-master ([pr#26515](#), Abhishek Lekshmanan)
- build/ops: cmake,do\_freebsd.sh: disable rdma features ([pr#22752](#), Kefu Chai)
- build/ops: cmake/modules/BuildDPDK.cmake: Build required DPDK libraries ([issue#36341](#), [pr#24487](#), Brad Hubbard)
- build/ops: cmake/modules/BuildRocksDB.cmake: enable compressions for rocksdb ([issue#24025](#), [pr#22181](#), Kefu Chai)
- build/ops: cmake,rgw: make amqp support optional ([pr#26555](#), Kefu Chai)
- build/ops: cmake,rpm,deb: install mgr plugins into /usr/share/ceph/mgr ([pr#26446](#), Kefu Chai)
- build/ops: cmake,seastar: pick up latest seastar ([pr#25474](#), Kefu Chai)
- build/ops,common: compressor: Fix build of Brotli Compressor ([pr#24967](#), BI SHUN KE)
- build/ops,common,core: test: make readable.sh fail if it doesn't run anything ([pr#24812](#), Greg Farnum)
- build/ops,core: cmake,common,filestore: silence gcc-8 warnings/errors ([pr#21837](#), Kefu Chai)
- build/ops,core,rbd: include/memory.h: remove memory.h ([pr#22690](#), Kefu Chai)
- build/ops,core: systemd: only restart 3 times in 30 minutes, as fast as possible ([issue#24368](#), [pr#22349](#), Greg Farnum)
- build/ops,core,tests: objectstore/test/fio: Fixed fio compilation when tcmalloc is used ([pr#23962](#), Adam Kupczyk)
- build/ops: credits.sh: Ignore package-lock.json and .xlf files ([pr#24762](#), Tiago Melo)
- build/ops: deb: drop redundant ceph-common recommends ([pr#20133](#), Nathan Cutler)
- build/ops: debian/control: change Architecture python plugins to "all" ([pr#26377](#), Kefu Chai)
- build/ops: debian/control: require fuse for ceph-fuse ([issue#21057](#), [pr#23675](#), Thomas Serlin)

- build/ops: debian: correct ceph-common relationship with older radosgw package ([pr#24996](#), Matthew Vernon)
- build/ops: debian: drop '-DUSE\_CRYPTOPP=OFF' from cmake options ([pr#22471](#), Kefu Chai)
- build/ops: debian: librados-dev should replace librados2-dev ([pr#25916](#), Kefu Chai)
- build/ops: debian/rules: fix ceph-mgr .pyc files left behind ([issue#26883](#), [pr#23615](#), Dan Mick)
- build/ops: deb,rpm,do\_cmake: switch to cmake3 ([pr#22896](#), Kefu Chai)
- build/ops: dmclock, cmake: sync up with ceph/dmclock, dmclock related cleanups ([issue#26998](#), [pr#23643](#), Kefu Chai)
- build/ops: dmclock: update dmclock submodule sha1 to tip of ceph/dmclock.git master ([pr#23837](#), Ricardo Dias)
- build/ops: do\_cmake.sh: automate py3 build options for certain distros ([pr#25205](#), Nathan Cutler)
- build/ops: do\_cmake.sh: SUSE builds need WITH\_RADOSGW\_AMQP\_ENDPOINT=OFF ([pr#26695](#), Nathan Cutler)
- build/ops: do\_freebsd.sh: FreeBSD building needs the llvm linker ([pr#25247](#), Willem Jan Withagen)
- build/ops: dout: declare dpp using decltype(auto) instead of auto ([pr#22207](#), Kefu Chai)
- build/ops: dpdk: drop dpdk submodule ([issue#24032](#), [pr#21856](#), Kefu Chai)
- build/ops: examples/Makefile: add -Wno-unused-parameter to avoid compile error ([pr#23581](#), You Ji)
- build/ops: Improving make check reliability ([pr#22441](#), Erwan Velu)
- build/ops: include: define errnos if not defined for better portability ([pr#25302](#), Willem Jan Withagen)
- build/ops: install-deps: check the exit status for the \$builddepcmd ([pr#22682](#), Yunchuan Wen)
- build/ops: install-deps: do not specify unknown options ([pr#24315](#), Kefu Chai)
- build/ops: install-deps: install setuptools before upgrading virtualenv ([pr#25039](#), Kefu Chai)
- build/ops: install-deps: nuke wheelhouse if it's stale ([pr#22028](#), Kefu Chai)

- build/ops: install-deps, run-make-check: use ceph-libboost repo ([issue#25186](#), [pr#23995](#), Kefu Chai)
- build/ops: install-deps.sh: Add Kerberos requirement for FreeBSD ([pr#25688](#), Willem Jan Withagen)
- build/ops: install-deps.sh: disable centos-sclo-rh-source ([issue#37707](#), [pr#25629](#), Brad Hubbard)
- build/ops: install-deps.sh: fix gcc detection and install pre-built libboost on bionic ([pr#25169](#), Changcheng Liu, Kefu Chai)
- build/ops: install-deps.sh: fix installing gcc on ubuntu when no old compiler ([pr#22488](#), Tomasz Setkowski)
- build/ops: install-deps.sh: import ubuntu-toolchain-r's key without keyserver ([pr#22964](#), Kefu Chai)
- build/ops: install-deps.sh: install libtool-ltdl-devel for building python-saml ([pr#25071](#), Kefu Chai)
- build/ops: install-deps.sh: refrain from installing/using lsb\_release, and other cleanup ([issue#18163](#), [pr#23361](#), Nathan Cutler)
- build/ops: install-deps.sh: Remove CR repo ([issue#13997](#), [pr#25211](#), Brad Hubbard, Alfredo Deza)
- build/ops: install-deps.sh: selectively install dependencies ([pr#26402](#), Kefu Chai)
- build/ops: install-deps.sh: set with\_seastar ([pr#23079](#), Nathan Cutler)
- build/ops: install-deps.sh: support install gcc7 in xenial aarch64 ([pr#22451](#), Yunchuan Wen)
- build/ops: install-deps.sh: Update python requirements for FreeBSD ([pr#25245](#), Willem Jan Withagen)
- build/ops: install-deps.sh: use the latest setuptools ([pr#26156](#), Kefu Chai)
- build/ops: install-deps: s/openldap-client/openldap24-client/ ([pr#23912](#), Kefu Chai)
- build/ops: libradosstriper: conditional compile ([pr#21983](#), Jesse Williamson)
- build/ops: make-debs.sh: clean dir to allow building deb packages multiple times ([pr#25177](#), Changcheng Liu)
- build/ops: man: skip directive starting with “..” ([pr#23580](#), Kefu Chai)
- build/ops,mgr: build: mgr: check for python's ssl version linkage ([issue#24282](#), [pr#22659](#), Kefu Chai, Abhishek Lekshmanan)

- build/ops,mgr: cmake,deb,rpm: remove cython 0.29's subinterpreter check, re-enable build with cython 0.29+ ([pr#25585](#), Tim Serong)
- build/ops: mgr/dashboard: Add html-linter ([pr#24273](#), Tiago Melo)
- build/ops: mgr/dashboard: Add i18n validation script ([pr#25179](#), Tiago Melo)
- build/ops: mgr/dashboard: Add package-lock.json ([pr#23285](#), Tiago Melo)
- build/ops: mgr/dashboard: Disable showing xi18n's progress ([pr#25427](#), Tiago Melo)
- build/ops: mgr/dashboard: Fix run-frontend-e2e-tests.sh ([pr#25157](#), Tiago Melo)
- build/ops: mgr/dashboard: fix the version of all frontend dependencies ([pr#22712](#), Tiago Melo)
- build/ops: mgr/dashboard: Remove angular build progress logs during cmake ([pr#23115](#), Tiago Melo)
- build/ops: mgr/dashboard: Update Node.js to current LTS ([pr#24932](#), Tiago Melo)
- build/ops: mgr/dashboard: Update node version ([pr#22639](#), Tiago Melo)
- build/ops: mgr/diskprediction: Replace local predictor model file ([pr#24484](#), Rick Chen)
- build/ops,mgr: mgr/dashboard: Fix building under FreeBSD ([pr#22562](#), Willem Jan Withagen)
- build/ops: move dmclock subtree into submodule ([pr#21651](#), Danny Al-Gaaf)
- build/ops,pybind: ceph: do not raise StopIteration within generator ([pr#25400](#), Jason Dillaman)
- build/ops,rbd: osd,mon,pybind: Make able to compile with Clang ([pr#21861](#), Adam C. Emerson)
- build/ops,rbd: selinux: add support for ceph iscsi ([pr#24936](#), Mike Christie)
- build/ops,rbd: systemd: enable ceph-rbd-mirror.target ([pr#24935](#), Sébastien Han)
- build/ops,rgw: build/rgw: unittest\_rgw\_dmclock\_scheduler does not need Boost\_LIBRARIES ([pr#26799](#), Willem Jan Withagen)
- build/ops,rgw: cls: build cls\_otp only WITH\_RADOSGW ([pr#22548](#), Piotr Dałek)
- build/ops,rgw: deb,rpm: package librgw\_admin\_user.{h,so.\*} ([pr#22205](#), Kefu Chai)
- build/ops: rocksdb: sync with upstream ([issue#23653](#), [pr#22236](#), Kefu Chai)
- build/ops: rpm: bump up required GCC version to 7.3.1 ([pr#24130](#), Kefu Chai)
- build/ops: rpm,deb: remove python-jinja2 dependency ([pr#26379](#), Kefu Chai)

- build/ops: rpm: do not exclude s390x build on openSUSE ([pr#26268](#), Nathan Cutler)
- build/ops: rpm: Fix Fedora error “No matching package to install: ‘Cython3’” ([issue#35831](#), [pr#23993](#), Brad Hubbard)
- build/ops: rpm: fix libradospp-devel runtime dependency ([pr#25491](#), Nathan Cutler)
- build/ops: rpm: fix seastar build dependencies for SUSE ([pr#23089](#), Nathan Cutler)
- build/ops: rpm: fix seastar build dependencies ([pr#23386](#), Nathan Cutler)
- build/ops: rpm: fix xmlsec1 build dependency for dashboard make check ([pr#26119](#), Nathan Cutler)
- build/ops: rpm: Install python2-Cython on f28 ([pr#26756](#), Brad Hubbard)
- build/ops: rpm: make ceph-grafana-dashboards own its directories ([issue#37485](#), [pr#25347](#), Nathan Cutler, Tim Serong)
- build/ops: rpm: make Python dependencies somewhat less confusing ([pr#25963](#), Nathan Cutler)
- build/ops: rpm: make sudo a build dependency ([pr#23077](#), Nathan Cutler)
- build/ops: rpm: package crypto libraries for all archs ([pr#26202](#), Nathan Cutler)
- build/ops: rpm: Package grafana dashboards ([pr#24735](#), Boris Ranto)
- build/ops: rpm: provide files moved from ceph-test ... ([issue#22558](#), [pr#20401](#), Nathan Cutler)
- build/ops: rpm: RHEL 8 fixes ([pr#26520](#), Ken Dreyer)
- build/ops: rpm: RHEL 8 needs Python 3 build ([pr#25223](#), Nathan Cutler)
- build/ops: rpm: stop install-deps.sh clobbering spec file Python build setting ([issue#37301](#), [pr#25181](#), Nathan Cutler, Brad Hubbard)
- build/ops: rpm: Use hardened LDFLAGS ([issue#36316](#), [pr#24425](#), Boris Ranto)
- build/ops: rpm: use updated gperftools ([issue#35969](#), [pr#24124](#), Kefu Chai)
- build/ops: rpm: Use updated gperftools-libs at runtime ([issue#36508](#), [pr#24652](#), Brad Hubbard)
- build/ops: run-make-check: enable -with-seastar option ([pr#22809](#), Kefu Chai)
- build/ops: run-make-check: set WITH\_SEASTAR with a non-empty string ([pr#23108](#), Kefu Chai)
- build/ops: run-make-check.sh: Adding ccache tuning for the CI ([pr#22847](#), Erwan Velu)

- build/ops: run-make-check.sh: ccache goodness for everyone ([issue#24817](#), [issue#24777](#), [pr#22867](#), Nathan Cutler)
- build/ops: run-make-check: should use sudo for running sysctl ([pr#23708](#), Kefu Chai)
- build/ops: run-make-check: Showing configuration before the build ([pr#23609](#), Erwan Velu)
- build/ops: seastar: lower the required yaml-cpp version to 0.5.1 ([pr#23255](#), Kefu Chai)
- build/ops: seastar: pickup the change to link pthread ([pr#25671](#), Kefu Chai)
- build/ops: selinux: Allow ceph to execute ldconfig ([pr#20118](#), Boris Ranto)
- build/ops: spdk: update to latest spdk-18.05 branch ([pr#22547](#), Kefu Chai)
- build/ops: spec: requires ceph base instead of common ([issue#37620](#), [pr#25503](#), Sébastien Han)
- build/ops: test: move ceph-dencoder to src/tools ([pr#23228](#), Kefu Chai)
- build/ops: test,qa: s/.libs/lib/ ([pr#20734](#), Kefu Chai)
- build/ops,tests: cmake,run-make-check: always enable WITH\_GTEST\_PARALLEL ([pr#23382](#), Kefu Chai)
- build/ops,tests: deb,rpm,qa: split dashboard package ([pr#26380](#), Kefu Chai)
- build/ops,tests: mgr/dashboard: Fix localStorage problem in Jest ([pr#23281](#), Tiago Melo)
- build/ops,tests: mgr/dashboard: Object Gateway user configuration ([pr#25494](#), Laura Paduano)
- build/ops,tests: src/test: Using gtest-parallel to speedup unittests ([pr#22577](#), Kefu Chai, Erwan Velu)
- build/ops,tests: tests/fio: fix build failures and ensure this is covered by run-make-check.sh ([pr#23231](#), Kefu Chai, Igor Fedotov)
- build/ops,tests: tests/qa: Adding \$ distro mix - rgw ([pr#21932](#), Yuri Weinstein)
- build/ops,tests: tools/ceph-dencoder: conditionally link against mds ([pr#25255](#), Kefu Chai)
- build/ops,tools: tool: link rbd-ggate against librados-cxx ([pr#24901](#), Willem Jan Withagen)
- ceph-disk: get\_partition\_dev() should fail until get\_dev\_path(partnam... ([pr#21415](#), Erwan Velu)

- cephfs: doc/releases: update CephFS mimic notes ([issue#23775](#), [pr#22232](#), Patrick Donnelly)
- cephfs: mgr/dashboard: NFS Ganesha management REST API ([pr#25918](#), Lenz Grimmer, Ricardo Dias, Jeff Layton)
- cephfs,mgr,pybind: pybind/mgr: Unified bits of volumes and orchestrator ([pr#25492](#), Sebastian Wagner)
- cephfs,mon: MDSMonitor: silence unable to load metadata ([pr#25693](#), Song Shun)
- cephfs,mon: mon/MDSMonitor: do not send redundant MDS health messages to cluster log ([issue#24308](#), [pr#22252](#), Sage Weil)
- cephfs: qa: fix symlink ([pr#23997](#), Patrick Donnelly)
- cephfs,rbd: osdc: Fix the wrong BufferHead offset ([pr#22495](#), dongdong tao)
- cephfs,rbd: osdc: optimize the code doing the BufferHead mapping ([pr#22509](#), dongdong tao)
- cephfs,rbd: osdc: reduce ObjectCacher's memory fragments ([issue#36192](#), [pr#24297](#), "Yan, Zheng")
- cephfs,tests: qa: fix run call args ([issue#36450](#), [pr#24597](#), Patrick Donnelly)
- cephfs,tests: qa: install python3-cephfs for fs suite ([pr#23411](#), Kefu Chai)
- cephfs,tests: qa/suites/powercycle: whitelist MDS\_SLOW\_REQUEST ([pr#23151](#), Neha Ojha)
- cephfs,tests: qa/workunits/suites/pjd.sh: use correct dir name ([pr#22233](#), Neha Ojha)
- ceph-volume: activate option --auto-detect-objectstore respects --no-systemd ([issue#36249](#), [pr#24355](#), Alfredo Deza)
- ceph-volume: Adapt code to support Python3 ([pr#25324](#), Volker Theile)
- ceph-volume: add --all flag to simple activate ([pr#26225](#), Jan Fajerski)
- ceph-volume add a \_\_release\_\_ string, to help version-conditional calls ([issue#25171](#), [pr#23332](#), Alfredo Deza)
- ceph-volume: add inventory command ([issue#24972](#), [pr#24859](#), Jan Fajerski)
- ceph-volume: Additional work on ceph-volume to add some choose\_disk capabilities ([issue#36446](#), [pr#24504](#), Erwan Velu)
- ceph-volume add new ceph-handlers role from ceph-ansible ([issue#36251](#), [pr#24336](#), Alfredo Deza)

- ceph-volume: adds a -prepare flag to lvm batch ([issue#36363](#), [pr#24587](#), Andrew Schoen)
- ceph-volume: add space between words ([pr#26246](#), Sébastien Han)
- ceph-volume: adds test for ceph-volume lvm list /dev/sda ([issue#24784](#), [issue#24957](#), [pr#23348](#), Andrew Schoen)
- ceph-volume: Add unit test ([pr#25321](#), Volker Theile)
- ceph-volume: allow to specify -cluster-fsid instead of reading from ceph.conf ([issue#26953](#), [pr#24407](#), Alfredo Deza)
- ceph-volume: an OSD ID must be exist and be destroyed before reuse ([pr#23093](#), Andrew Schoen, Ron Allred)
- ceph-volume batch: allow journal+block.db sizing on the CLI ([issue#36088](#), [pr#24201](#), Alfredo Deza)
- ceph-volume batch: allow -osds-per-device, default it to 1 ([issue#35913](#), [pr#24060](#), Alfredo Deza)
- ceph-volume batch carve out lvs for bluestore ([pr#24019](#), Alfredo Deza)
- ceph-volume batch command ([issue#24492](#), [pr#23075](#), Alfredo Deza)
- ceph-volume: batch tests for mixed-type of devices ([issue#35535](#), [issue#27210](#), [pr#23963](#), Alfredo Deza)
- ceph-volume custom cluster names fail on filestore trigger ([issue#27210](#), [pr#24251](#), Alfredo Deza)
- ceph-volume: do not pin the testinfra version for the simple tests ([pr#23268](#), Andrew Schoen)
- ceph-volume: do not send (lvm) stderr/stdout to the terminal, use the logfile ([issue#36492](#), [pr#24738](#), Alfredo Deza)
- ceph-volume do not use stdin in luminous ([issue#25173](#), [pr#23355](#), Alfredo Deza)
- ceph-volume: don't create osd['block.db'] by default ([issue#38472](#), [pr#26627](#), Jan Fajerski)
- ceph-volume: earlier detection for -journal and -filestore flag requirements ([issue#24794](#), [pr#24150](#), Alfredo Deza)
- ceph-volume: enable device discards ([issue#36532](#), [pr#24676](#), Jonas Jelten)
- ceph-volume enable -no-systemd flag for simple sub-command ([issue#36470](#), [pr#24998](#), Alfredo Deza)
- ceph-volume: enable the ceph-osd during lvm activation ([issue#24152](#), [pr#23321](#),

Dan van der Ster)

- ceph-volume ensure encoded bytes are always used ([issue#24993](#), [pr#23289](#), Alfredo Deza)
- ceph-volume: error on commands that need ceph.conf to operate ([issue#23941](#), [pr#22724](#), Andrew Schoen)
- ceph-volume expand auto engine for multiple devices on filestore ([issue#24553](#), [pr#23731](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: expand auto engine for single type devices on filestore ([issue#24960](#), [pr#23532](#), Alfredo Deza)
- ceph-volume expand on the LVM API to create multiple LVs at different sizes ([issue#24020](#), [pr#22426](#), Alfredo Deza)
- ceph-volume: extract flake8 config ([pr#24674](#), Mehdi Abaakouk)
- ceph-volume: fix Batch object in py3 environments ([pr#25203](#), Jan Fajerski)
- ceph-volume: fix journal and filestore data size in lvm batch -report ([issue#36242](#), [pr#24274](#), Andrew Schoen)
- ceph-volume: fix JSON output in inventory ([issue#37390](#), [pr#25224](#), Sebastian Wagner)
- ceph-volume: Fix TypeError: join() takes exactly one argument (2 given) ([issue#37595](#), [pr#25469](#), Sebastian Wagner)
- ceph-volume fix TypeError on dmcrypt when using Python3 ([pr#26034](#), Alfredo Deza)
- ceph-volume fix zap not working with LVs ([issue#35970](#), [pr#24077](#), Alfredo Deza)
- ceph-volume: implement \_\_format\_\_ in Size to format sizes in py3 ([issue#38291](#), [pr#26401](#), Jan Fajerski)
- ceph-volume initial take on auto sub-command ([pr#21803](#), Alfredo Deza)
- ceph-volume: introduce class hierarchy for strategies ([issue#37389](#), [pr#25238](#), Jan Fajerski)
- ceph-volume: lsblk can fail to find PARTLABEL, must fallback to blkid ([issue#36098](#), [pr#24330](#), Alfredo Deza)
- ceph-volume lvm.activate conditional mon-config on prime-osd-dir ([issue#25216](#), [pr#23375](#), Alfredo Deza)
- ceph-volume lvm.activate Do not search for a MON configuration ([pr#22393](#), Wido den Hollander)
- ceph-volume lvm batch allow extra flags (like dmcrypt) for bluestore

- ([issue#26862](#), [pr#23448](#), Alfredo Deza)
- ceph-volume lvm.batch remove non-existent sys\_api property ([issue#34310](#), [pr#23787](#), Alfredo Deza)
  - ceph-volume lvm.listing only include devices if they exist ([issue#24952](#), [pr#23129](#), Alfredo Deza)
  - ceph-volume lvm.prepare update help to indicate partitions are needed, not devices ([issue#24795](#), [pr#24394](#), Alfredo Deza)
  - ceph-volume: make Device hashable to allow set of Device list in py3 ([issue#38290](#), [pr#26399](#), Jan Fajerski)
  - ceph-volume: make lvm batch idempotent ([issue#26864](#), [pr#24404](#), Andrew Schoen)
  - ceph-volume: mark a device not available if it belongs to ceph-disk ([pr#26084](#), Andrew Schoen)
  - ceph-volume normalize comma to dot for string to int conversions ([issue#37442](#), [pr#25674](#), Alfredo Deza)
  - ceph-volume: patch Device when testing ([issue#36768](#), [pr#25063](#), Alfredo Deza)
  - ceph-volume process.call with stdin in Python 3 fix ([issue#24993](#), [pr#23141](#), Alfredo Deza)
  - ceph-volume: provide a nice error message when missing ceph.conf ([pr#22828](#), Andrew Schoen)
  - ceph-volume: PVolumes.get() should return one PV when using name or uuid ([issue#24784](#), [pr#23234](#), Andrew Schoen)
  - ceph-volume: refuse to zap mapper devices ([issue#24504](#), [pr#22764](#), Andrew Schoen)
  - ceph-volume: reject devices that have existing GPT headers ([issue#27062](#), [pr#25098](#), Andrew Schoen)
  - ceph-volume: remove iteritems instances ([issue#38299](#), [pr#26403](#), Jan Fajerski)
  - ceph-volume: remove LVs when using zap -destroy ([pr#25093](#), Alfredo Deza)
  - ceph-volume remove version reporting from help menu ([issue#36386](#), [pr#24531](#), Alfredo Deza)
  - ceph-volume: rename Device property valid to available ([issue#36701](#), [pr#25007](#), Jan Fajerski)
  - ceph-volume: replace testinfra command with py.test ([issue#38568](#), [pr#26739](#), Alfredo Deza)
  - ceph-volume: Restore SELinux context ([pr#23278](#), Boris Ranto)

- ceph-volume: revert partition as disk ([issue#37506](#), [pr#25390](#), Jan Fajerski)
- ceph-volume: run tests without waiting on ceph repos ([pr#23697](#), Andrew Schoen)
- ceph-volume: set number of osd ports in the tests ([pr#26753](#), Andrew Schoen)
- ceph-volume: set permissions right before prime-osd-dir ([issue#37486](#), [pr#25477](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: simple scan will now scan all running ceph-disk OSDs ([pr#26826](#), Andrew Schoen)
- ceph-volume: skip processing devices that don't exist when scanning system disks ([issue#36247](#), [pr#24372](#), Alfredo Deza)
- ceph-volume: sort and align lvm list output ([pr#21812](#), Theofilos Mouratidis)
- ceph-volume systemd import main so console\_scripts work for executable ([issue#36648](#), [pr#24840](#), Alfredo Deza)
- ceph-volume tests destroy osds on monitor hosts ([pr#22437](#), Alfredo Deza)
- ceph-volume tests do not include admin keyring in OSD nodes ([issue#24417](#), [pr#22399](#), Alfredo Deza)
- ceph-volume tests/functional add mgrs daemons to lvm tests ([issue#26879](#), [pr#23489](#), Alfredo Deza)
- ceph-volume tests.functional add notario dep for ceph-ansible ([pr#22116](#), Alfredo Deza)
- ceph-volume tests/functional declare ceph-ansible roles instead of importing them ([issue#37805](#), [pr#25820](#), Alfredo Deza)
- ceph-volume tests.functional fix typo when stopping osd.0 in filestore ([issue#37675](#), [pr#25594](#), Alfredo Deza)
- ceph-volume: tests.functional inherit SSH\_ARGS from ansible ([issue#34311](#), [pr#23788](#), Alfredo Deza)
- ceph-volume tests/functional run lvm list after OSD provisioning ([issue#24961](#), [pr#23116](#), Alfredo Deza)
- ceph-volume tests/functional use Ansible 2.6 ([pr#23182](#), Alfredo Deza)
- ceph-volume tests install ceph-ansible's requirements.txt dependencies ([issue#36672](#), [pr#24881](#), Alfredo Deza)
- ceph-volume tests patch \_\_release\_\_ to mimic always for stdin keys ([pr#23398](#), Alfredo Deza)
- ceph-volume tests.systemd update imports for systemd module ([issue#36704](#),

pr#24937, Alfredo Deza)

- ceph-volume: test with multiple NVME drives ([issue#37409](#), [pr#25354](#), Andrew Schoen)
- ceph-volume: unmount lvs correctly before zapping ([issue#24796](#), [pr#23117](#), Andrew Schoen)
- ceph-volume: update testing playbook ‘deploy.yml’ ([pr#26397](#), Guillaume Abrioux)
- ceph-volume: update version of ansible to 2.6.x for simple tests ([pr#23263](#), Andrew Schoen)
- ceph-volume: use console\_scripts ([issue#36601](#), [pr#24773](#), Mehdi Abaakouk)
- ceph-volume: use our own testinfra suite for functional testing ([pr#26685](#), Andrew Schoen)
- ceph-volume util.encryption don’t push stderr to terminal ([issue#36246](#), [pr#24399](#), Alfredo Deza)
- ceph-volume util.encryption robust blkid+lsblk detection of lockbox ([pr#24977](#), Alfredo Deza)
- ceph-volume zap devices associated with an OSD ID and/or OSD FSID ([pr#25429](#), Alfredo Deza)
- ceph-volume zap: improve zapping to remove all partitions and all LVs, encrypted or not ([issue#37449](#), [pr#25330](#), Alfredo Deza)
- cleanup: Clean up warnings ([pr#23919](#), Adam C. Emerson)
- cli: dump osd-fsid as part of osd find <id> ([pr#26015](#), Noah Watkins)
- cmake: add “add\_npm\_command()” command ([pr#22636](#), Kefu Chai)
- cmake: Add cls\_opt for vstart target ([pr#22538](#), Ali Maredia)
- cmake: add dpdk::dpdk if dpdk is built or found ([issue#24948](#), [pr#23620](#), Nathan Cutler, Kefu Chai)
- cmake: add option WITH\_LIBRADOSSTRIPER ([pr#23732](#), Kefu Chai)
- cmake: allow setting of the CTest timeout during building ([pr#22800](#), Willem Jan Withagen)
- cmake: always prefer local symbols ([issue#25154](#), [pr#23320](#), Kefu Chai)
- cmake: always turn off bjam debugging output ([pr#22204](#), Kefu Chai)
- cmake: bump up the required boost version to 1.67 ([pr#22392](#), Kefu Chai)

- cmake: bump up the required fmt version ([pr#23283](#), Kefu Chai)
- cmake: cleanups ([pr#23166](#), Kefu Chai)
- cmake: cleanups ([pr#23279](#), Kefu Chai)
- cmake: cleanups ([pr#23300](#), Kefu Chai)
- cmake,crimson/net: add keepalive support, and enable unittest\_seastar\_messenger in "make check" ([pr#23642](#), Kefu Chai)
- cmake: detect armv8 crc and crypto feature using CHECK\_C\_COMPILER\_FLAG ([issue#17516](#), [pr#24168](#), Kefu Chai)
- cmake: disable -Werror-stringop-truncation for rocksdb ([pr#22591](#), Kefu Chai)
- cmake: do not check for aligned\_alloc() anymore ([issue#23653](#), [pr#22046](#), Kefu Chai)
- cmake: do not depend on \${DPDK\_LIBRARIES} if not using bundled dpdk ([issue#24449](#), [pr#22938](#), Kefu Chai)
- cmake: do not install hello demo module ([pr#21886](#), John Spray)
- cmake: do not link against common\_crc\_aarch64 ([pr#23366](#), Kefu Chai)
- cmake: do not pass -B{symbolic,symbolic-functions} to linker on FreeBSD ([pr#24920](#), Willem Jan Withagen)
- cmake: do not pass unnecessary param to setup.py ([pr#25186](#), Kefu Chai)
- cmake: do not use Findfmt.cmake for checking libfmt-dev ([pr#23390](#), Kefu Chai)
- cmake: do not use plain target\_link\_libraries(rgw\_a ...) ([pr#24515](#), Kefu Chai)
- cmake: enable RTTI for both debug and release RocksDB builds ([pr#22286](#), Igor Fedotov)
- cmake: find a python2 interpreter for gtest-parallel ([pr#22931](#), Kefu Chai)
- cmake: find liboath using the correct name ([pr#22430](#), Kefu Chai)
- cmake: fix a cmake error when with -DALLOCATOR=jemalloc ([pr#23380](#), Jianpeng Ma)
- cmake: fix build WITH\_SYSTEM\_BOOST=ON ([pr#23510](#), Kefu Chai)
- cmake: fix compilation with distcc and other compiler wrappers ([pr#24605](#), Alexey Sheplyakov, Kefu Chai)
- cmake: fix cython target in test/CMakeFile.txt ([pr#22295](#), Jan Fajerski)
- cmake: fix Debug build WITH\_SEASTAR=ON ([pr#23567](#), Kefu Chai)

- cmake: fixes to enable WITH\_ASAN with clang and GCC ([pr#24692](#), Kefu Chai)
- cmake: fix find system rockdb ([pr#22439](#), Alexey Shabalin)
- cmake: fix std::filesystem detection and extract sanitizer detection into its own module ([pr#23384](#), Kefu Chai)
- cmake: fix syntax error of set() ([pr#26582](#), Kefu Chai)
- cmake: fix the build WITH\_DPDK=ON ([pr#23650](#), Kefu Chai, Casey Bodley)
- cmake: fix version matching for Findfmt ([pr#23996](#), Mohamad Gebai)
- cmake: fix "WITH\_STATIC\_LIBSTDCXX" ([pr#22990](#), Kefu Chai)
- cmake: let rbd\_api depend on librbd-tp ([pr#25641](#), Kefu Chai)
- cmake: link against gtest in a better way ([pr#23628](#), Kefu Chai)
- cmake: link ceph-osd with common statically ([pr#22720](#), Radoslaw Zarzynski)
- cmake: link compressor plugins against lib the modern way ([pr#23852](#), Kefu Chai)
- cmake: make -DWITH\_MGR=OFF work ([pr#22077](#), Jianpeng Ma)
- cmake: Make the tests for finding Filesystem with more serious functions ([pr#26316](#), Willem Jan Withagen)
- cmake: modularize src/perfglue ([pr#23254](#), Kefu Chai)
- cmake: move ceph-osdomap-tool, ceph-monstore-tool out of ceph-test ([pr#19964](#), runsi)
- cmake: move crypto\_plugins target ([pr#21891](#), Casey Bodley)
- cmake: no libradosstriper headers if WITH\_LIBRADOSSTRIPER=OFF ([issue#35922](#), [pr#24029](#), Nathan Cutler, Kefu Chai)
- cmake: no need to add "-D" before definitions ([pr#23795](#), Kefu Chai)
- cmake: oath lives in liboath ([pr#22494](#), Willem Jan Withagen)
- cmake: only build extra boost libraries only if WITH\_SEASTAR ([pr#22521](#), Kefu Chai)
- cmake: remove checking for GCC 5.1 ([pr#24477](#), Kefu Chai)
- cmake: remove deleted rgw\_request.cc from CMakeLists.txt ([pr#22186](#), Casey Bodley)
- cmake: Remove embedded 'cephd' code ([pr#21940](#), Dan Mick)
- cmake: remove workarounds for supporting cmake 2.x ([pr#22912](#), Kefu Chai)

- cmake: rgw\_common should depend on tracing headers ([pr#22367](#), Kefu Chai)
- cmake: rocksdb related cleanup ([pr#23441](#), Kefu Chai)
- cmake: should link against libatomic if libcxx/libstdc++ does not off... ([pr#22952](#), Kefu Chai)
- cmake: update fio version from 3.5 to 540e235cd276e63c57 ([pr#22019](#), Jianpeng Ma)
- cmake: use \$CMAKE\_BINARY\_DIR for default \$CEPH\_BUILD\_VIRTUALENV ([issue#36737](#), [pr#26091](#), Kefu Chai)
- cmake: use javac -h for creating JNI native headers ([issue#24012](#), [pr#21822](#), Kefu Chai)
- cmake: use OpenSSL::Crypto instead of OPENSSL\_LIBRARIES ([pr#24368](#), Kefu Chai)
- cmake: vstart target can build WITH\_CEPHFS/RBD/MGR=OFF ([pr#25204](#), Casey Bodley)
- common: add adaptor for seastar::temporary\_buffer ([pr#22454](#), Kefu Chai, Casey Bodley)
- common: add a generic async Completion for use with boost::asio ([pr#21914](#), Casey Bodley)
- common: add lockless md\_config\_t ([pr#22710](#), Kefu Chai)
- common: async/dpdk: when enable dpdk, multiple message queue defect ([pr#25404](#), zhangyongsheng)
- common: auth/cephx: minor code cleanup ([pr#21155](#), runsisi)
- common: auth, common: cleanups ([pr#26383](#), Kefu Chai)
- common: auth,common: use ceph::mutex instead of LockMutex ([pr#24263](#), Kefu Chai)
- common: avoid the overhead of `__ANNOTATE_HAPPENS__*` in NDEBUG builds ([pr#25129](#), Radoslaw Zarzynski)
- common: be more informative if set PID-file fails ([pr#23647](#), Willem Jan Withagen)
- common: blkdev: Rework API and add FreeBSD support ([pr#24658](#), Alan Somers)
- common: buffer: mark the iterator traits “public” ([pr#25409](#), Kefu Chai)
- common: calculate stddev on the fly ([pr#21461](#), Yao Zongyou)
- common: ceph.in: use correct module for cmd flags ([pr#26454](#), Patrick Donnelly)
- common: ceph-volume add device\_id to inventory listing ([pr#25201](#), Jan Fajerski)
- common: changes to address FTBFS on fc30 ([pr#26301](#), Kefu Chai)

- common: common/admin\_socket: add new api unregister\_commands(AdminSocketHook ... ([pr#21718](#), Jianpeng Ma))
- common: common,auth,crimson: add logging to crimson ([pr#23957](#), Kefu Chai)
- common: common/buffer: fix compiler bug when enable DEBUG\_BUFFER ([pr#25848](#), Jianpeng Ma)
- common: common/buffer: remove repeated condition-check ([pr#25420](#), Jianpeng Ma)
- common: common/config: add ConfigProxy for crimson ([pr#23074](#), Kefu Chai)
- common: common/config: fix the lock in ConfigProxy::diff() ([pr#23276](#), Kefu Chai)
- common: common/config\_values: friend md\_config\_impl<> ([pr#23020](#), Mykola Golub, Kefu Chai)
- common: common: drop the unused methods from SharedLRU ([pr#26224](#), Radoslaw Zarzynski)
- common: common/KeyValueDB: Get rid of validate parameter ([pr#25377](#), Adam Kupczyk)
- common: common/numa: Add shim routines for NUMA on FreeBSD ([pr#25920](#), Willem Jan Withagen)
- common: common, osd: set mclock priority as 1 by default ([pr#26022](#), Abhishek Lekshmanan)
- common: common/random\_cache: remove unused RandomCache ([pr#26253](#), Kefu Chai)
- common: common/shared\_cache: add lockless SharedLRU ([pr#22736](#), Kefu Chai)
- common: common/shared\_cache: bumps it to the front of the LRU if key existed ([pr#25370](#), Jianpeng Ma)
- common: common/shared\_cache: fix racing issues ([pr#25150](#), Jianpeng Ma)
- common: common/util: pass real hostname when running in kubernetes/rook container ([pr#23798](#), Sage Weil)
- common: complete all throttle blockers when we set average or max to 0 ([issue#36715](#), [pr#24965](#), Dongsheng Yang)
- common,core: msg/async: clean up local buffers on dispatch ([issue#35987](#), [pr#24111](#), Greg Farnum)
- common,core,tests: qa/tests: update links for centos latest to point to 7.5 ([pr#22923](#), Vasu Kulkarni)
- common/crc/aarch64: Added cpu feature pmull and make aarch64 specific... ([pr#22178](#), Adam Kupczyk)

- common: crimson/common: write configs synchronously on shard.0 ([pr#23284](#), Kefu Chai)
- common,crimson: port perfcounters to seastar ([pr#24141](#), chunmei Liu)
- common: crypto: QAT based Encryption for RGW ([pr#19386](#), Ganesh Maharaj Mahalingam)
- common: crypto: use ceph\_assert\_always for assertions ([pr#23654](#), Casey Bodley)
- common: define BOOST\_COROUTINES\_NO\_DEPRECATED\_WARNING if not yet ([pr#26502](#), Kefu Chai)
- common: drop allocation tracking from bufferlist ([pr#25454](#), Radoslaw Zarzynski)
- common: drop append\_buffer from bufferlist. Use simple carriage instead ([pr#25077](#), Radoslaw Zarzynski)
- common: drop at\_buffer\_{head,tail} from buffer::ptr ([pr#25422](#), Radoslaw Zarzynski)
- common: drop/mark-as-final getters of buffer::raw for palign ([pr#24087](#), Radoslaw Zarzynski)
- common: drop static\_assert.h as it looks unused ([pr#22743](#), Radoslaw Zarzynski)
- common: drop the unused buffer::raw\_mmap\_pages ([pr#24040](#), Radoslaw Zarzynski)
- common: drop the unused zero-copy facilities in ceph::bufferlist ([pr#24031](#), Radoslaw Zarzynski)
- common: drop unused get\_max\_pipe\_size() in buffer.cc ([pr#25432](#), Radoslaw Zarzynski)
- common: ec: lrc doesn't depend on crosstalks between bufferlists anymore ([pr#25595](#), Radoslaw Zarzynski)
- common: expand meta in parse\_argv() ([pr#23474](#), Kefu Chai)
- common: fix access and add name for the token bucket throttle ([pr#25372](#), Shiyang Ruan)
- common: Fix Alpine compatibility for TEMP\_FAILURE\_RETRY and ACCESSPERMS ([pr#24813](#), Willem Jan Withagen)
- common: fix a racing in PerfCounters::perf\_counter\_data\_any\_d::read\_avg ([issue#25211](#), [pr#23362](#), ludehp)
- common: fix for broken rbdmap parameter parsing ([pr#24446](#), Marc Schoechlin)
- common: fix missing include boost/noncopyable.hpp ([pr#24278](#), Willem Jan Withagen)

- common: fix typo in rados bench write JSON output ([issue#24199](#), [pr#22112](#), Sandor Zeestraten)
- common: fix typos in BackoffThrottle ([pr#24691](#), Shiyang Ruan)
- common: Formatters: improve precision of double numbers ([pr#25745](#), Коренберг Марк)
- common: .gitignore: Ignore .idea directory ([pr#24237](#), Volker Theile)
- common: hint bufferlist's buffer\_track\_c\_str accordingly ([pr#25424](#), Radoslaw Zarzynski)
- common: hypercombined bufferlist ([pr#24882](#), Radoslaw Zarzynski)
- common: include/compat.h: make pthread\_get\_name\_np work when available ([pr#23641](#), Willem Jan Withagen)
- common: include/include/types.h early, otherwise Clang will error ([pr#22493](#), Willem Jan Withagen)
- common: include/types: move operator<< into the proper namespace ([pr#23767](#), Kefu Chai)
- common: include/types: space between number and units ([pr#22063](#), Sage Weil)
- common: librados,rpm,deb: various fixes to address librados3 transition and cleanups in librados ([pr#24896](#), Kefu Chai)
- common: make CEPH\_BUFFER\_ALLOC\_UNIT known at compile-time ([pr#26259](#), Radoslaw Zarzynski)
- common: mark BlkDev::serial() const to match with its declaration ([pr#24702](#), Willem Jan Withagen)
- common: messages: define HEAD\_VERSION and COMPAT\_VERSION inlined ([pr#23623](#), Kefu Chai)
- common,mgr: mgr/MgrClient: make some noise for a user if no mgr daemon is running ([pr#23492](#), Sage Weil)
- common: mon/MonClient: set configs via finisher ([issue#24118](#), [pr#21984](#), Sage Weil)
- common: msg/async: fix FTBFS of dpdk ([pr#23168](#), Kefu Chai)
- common: msg/async: Skip the duplicated processing of the same link ([pr#20952](#), shangfufei)
- common: msg/msg\_types.h: do not cast ceph\_entity\_name to entity\_name\_t for printing ([pr#26315](#), Kefu Chai)

- common: msgr/async/rdma: Return from poll system call with EINTR should be retried ([pr#25138](#), Stig Telfer)
- common: Mutex -> ceph::mutex ([issue#12614](#), [pr#25105](#), Kefu Chai, Sage Weil)
- common: optimize reference counting in bufferlist ([pr#25082](#), Radoslaw Zarzynski)
- common: OpTracker doesn't visit TrackedOp when nref == 0 ([issue#24037](#), [pr#22156](#), Radoslaw Zarzynski)
- common: os/filestore: fix throttle configurations ([pr#21926](#), Li Wang)
- common,performance: auth,common: add lockless auth ([pr#23591](#), Kefu Chai)
- common,performance: common/assert: mark assert helpers with [[gnu::cold]] ([pr#23326](#), Kefu Chai)
- common,performance: compressor: add QAT support ([pr#19714](#), Qiaowei Ren)
- common,performance: denc: fix internal fragmentation when decoding ptr in bl ([pr#25264](#), Kefu Chai)
- common,rbd: misc: mark constructors as explicit ([pr#21637](#), Danny Al-Gaaf)
- common: reinit StackStringStream on clear ([pr#25751](#), Patrick Donnelly)
- common: reintroduce async SharedMutex ([issue#24124](#), [pr#22698](#), Casey Bodley)
- common: Reverse deleted include ([pr#23838](#), Willem Jan Withagen)
- common: Revert "common: add an async SharedMutex" ([issue#24124](#), [pr#21986](#), Casey Bodley)
- common,rgw: cls/rbd: init local var with known value ([pr#25588](#), Kefu Chai)
- common,tests: run-standalone.sh: Need double-quotes to handle | in core\_pattern on all distributions ([issue#38325](#), [pr#26436](#), David Zafman)
- common,tests: test\_shared\_cache: fix memory leak ([pr#25215](#), Jianpeng Ma)
- common: vstart: do not attempt to re-initialize dashboard for existing cluster ([pr#23261](#), Jason Dillaman)
- core: Add support for osd\_delete\_sleep configuration value ([issue#36474](#), [pr#24749](#), David Zafman)
- core: auth: drop the RWLock in AuthClientHandler ([pr#23699](#), Kefu Chai)
- core: auth krb: Fix Kerberos build warnings ([pr#25639](#), Daniel Oliveira)
- core: build: disable kerberos for nautilus ([pr#26258](#), Sage Weil)
- core: ceph\_argparse: fix -verbose ([pr#25961](#), Patrick Nawracay)

- core: ceph.in: friendlier message on EPERM ([issue#25172](#), [pr#23330](#), John Spray)
- core: ceph.in: write bytes to stdout in raw\_write() ([pr#25280](#), Kefu Chai)
- core: ceph\_test\_rados\_api\_misc: remove obsolete LibRadosMiscPool.PoolCreationRace ([issue#24150](#), [pr#22042](#), Sage Weil)
- core: Clang misses <optional> include ([pr#23768](#), Willem Jan Withagen)
- core: common/blkdev.h: use std::string ([pr#25783](#), Neha Ojha)
- core: common/options: remove unused ms async affinity options ([pr#26099](#), Josh Durgin)
- core: common/util.cc: add CONTAINER\_NAME processing for metadata ([pr#25383](#), Dan Mick)
- core: compressor: building error for QAT decompress ([pr#22609](#), Qiaowei Ren)
- core: crush, osd: handle multiple parents properly when applying pg upmaps ([issue#23921](#), [pr#21815](#), xiexingguo)
- core: erasure-code: add clay codes ([issue#19278](#), [pr#24291](#), Myna V, Sage Weil)
- core: erasure-code: fixes alignment issue when clay code is used with jerasure, cauchy\_orig ([pr#24586](#), Myna)
- core: global/signal\_handler.cc: report assert\_file as correct name ([pr#23738](#), Dan Mick)
- core: include/rados: clarify which flags go where for copy\_from ([pr#24497](#), Ilya Dryomov)
- core: include/rados.h: hide CEPH\_OSDMAP\_PGLOG\_HARDLIMIT from ceph -s ([pr#25887](#), Neha Ojha)
- core: kv/KeyValueDB: Move PriCache implementation to ShardedCache ([pr#25925](#), Mark Nelson)
- core: kv/KeyValueDB: return const char\* from MergeOperator::name() ([issue#26875](#), [pr#23477](#), Sage Weil)
- core: messages/MOSDPGScan: fix initialization of query\_epoch ([pr#22408](#), wumingqiao)
- core: mgr/balancer: add cmd to list all plans ([issue#37418](#), [pr#21937](#), Yang Honggang)
- core: mgr/BaseMgrModule: drop GIL for ceph\_send\_command ([issue#38537](#), [pr#26723](#), Sage Weil)
- core: mgr/MgrClient: Protect daemon\_health\_metrics ([issue#23352](#), [pr#23404](#), Kjetil

Joergensen, Brad Hubbard)

- core,mgr: mon/MgrMonitor: change ‘unresponsive’ message to info level ([issue#24222](#), [pr#22158](#), Sage Weil)
- core,mgr,rbd: mgr: generalize osd perf query and make counters accessible from modules ([pr#25114](#), Mykola Golub)
- core,mgr,rbd: osd: support more dynamic perf query subkey types ([pr#25371](#), Mykola Golub)
- core,mgr,rbd,rgw: rgw, common: Fixes SCA issues ([pr#22007](#), Danny Al-Gaaf)
- core: mgr/smart: remove obsolete smart module ([pr#26411](#), Sage Weil)
- core: mon/LogMonitor: call no\_reply() on ignored log message ([pr#22098](#), Sage Weil)
- core: mon/MonClient: avoid using magic number for the MAuth::protocol ([pr#23747](#), Kefu Chai)
- core: mon/MonClient: extract MonSub out ([pr#23688](#), Kefu Chai)
- core: mon/MonClient: use scoped\_guard instead of goto ([pr#24304](#), Kefu Chai)
- core,mon: mon,osd: dump “compression\_algorithms” in “mon metadata” ([issue#22420](#), [pr#21809](#), Kefu Chai, Casey Bodley)
- core,mon: mon/OSDMonitor: no\_reply on MOSDFailure messages ([issue#24322](#), [pr#22259](#), Sage Weil)
- core,mon: mon/OSDMonitor: Warnings for expected\_num\_objects ([issue#24687](#), [pr#23072](#), Douglas Fuller)
- core: mon/OSDMonitor: two “ceph osd crush class rm” fixes ([pr#24657](#), xie xingguo)
- core: mon/PGMap: fix PGMapDigest decode ([pr#22066](#), Sage Weil)
- core: mon/PGMap: include unknown PGs in ‘pg ls’ ([pr#24032](#), Sage Weil)
- core: msg/async: do not trigger RESETSESSION from connect fault during connection phase ([issue#36612](#), [pr#25343](#), Sage Weil)
- core: msg/async/Event: clear time\_events on shutdown ([issue#24162](#), [pr#22093](#), Sage Weil)
- core: msg/async: fix banner\_v1 check in ProtocolV2 ([pr#26714](#), Yingxin Cheng)
- core: msg/async: fix include in frames\_v2.h ([pr#26711](#), Yingxin Cheng)
- core: msg/async: fix is\_queued() semantics ([pr#24693](#), Ilya Dryomov)

- core: msg/async: keep connection alive only actually sending ([pr#24301](#), Haomai Wang, Kefu Chai)
- core: os/bluestore: fix deep-scrub operation against disk silent errors ([pr#23629](#), Xiaoguang Wang)
- core: os/bluestore: fix flush\_commit locking ([issue#21480](#), [pr#22083](#), Sage Weil)
- core: OSD: add impl for filestore to get dbstatistics ([issue#24591](#), [pr#22633](#), lvshuhua)
- core: osdc: Change 'bool budgeted' to 'int budget' to avoid recalculating ([pr#21242](#), Jianpeng Ma)
- core: OSD: ceph-osd parent process need to restart log service after fork ([issue#24956](#), [pr#23090](#), redickwang)
- core: osdc/Objecter: fix split vs reconnect race ([issue#22544](#), [pr#23850](#), Sage Weil)
- core: osdc/Objecter: no need null pointer check for op->session anymore ([pr#25230](#), runsisi)
- core: osdc/Objecter: possible race condition with connection reset ([issue#36183](#), [pr#24276](#), Jason Dillaman)
- core: osdc: self-managed snapshot helper should catch decode exception ([issue#24000](#), [pr#21804](#), Jason Dillaman)
- core: osd, librados: add unset-manifest op ([pr#21999](#), Myoungwon Oh)
- core: osd,mds: make 'config rm ...' idempotent ([issue#24408](#), [pr#22395](#), Sage Weil)
- core: osd/mon: fix upgrades for pg log hard limit ([issue#36686](#), [pr#25816](#), Neha Ojha, Yuri Weinstein)
- core: osd,mon: increase mon\_max\_pg\_per\_osd to 250 ([pr#23251](#), Neha Ojha)
- core: osd,mon,msg: use intrusive\_ptr for holding Connection::priv ([issue#20924](#), [pr#22292](#), Kefu Chai)
- core: osd/OSD: choose heartbeat peers more carefully ([pr#23487](#), xie xingguo)
- core: osd/OSD: drop extra/wrong \*unregister\_pg\* ([pr#21816](#), xiexingguo)
- core: osd/OSDMap: be more aggressive when trying to balance ([issue#37940](#), [pr#26039](#), xie xingguo)
- core: osd/OSDMap: drop local pool filter in calc\_pg\_upmaps ([pr#26605](#), xie xingguo)
- core: osd/OSDMap: fix CEPHX\_V2 osd requirement to nautilus, not mimic ([pr#23249](#),

Sage Weil)

- core: osd/OSDMap: fix upmap mis-killing for erasure-coded PGs ([pr#25365](#), ningtao, xie xingguo)
- core: osd/OSDMap: potential access violation fix ([issue#37881](#), [pr#25930](#), xie xingguo)
- core: osd/OSDMap: using std::vector::reserve to reduce memory reallocation ([pr#26478](#), xie xingguo)
- core: osd/OSD: ping monitor if we are stuck at \_\_waiting\_for\_healthy\_\_ ([pr#23958](#), xie xingguo)
- core: osd/OSD: preallocate for \_\_get\_pgs/\_\_get\_pgids to avoid reallocate ([pr#25434](#), Jianpeng Ma)
- core: osd/PG: async-recovery should respect historical missing objects ([pr#24004](#), xie xingguo)
- core: osd/PG.cc: account for missing set irrespective of last\_complete ([issue#37919](#), [pr#26175](#), Neha Ojha)
- core: osd/PG: create new PGs from activate in last\_peering\_reset epoch ([issue#24452](#), [pr#22478](#), Sage Weil)
- core: osd/PG: do not choose stray osds as async\_recovery\_targets ([pr#22330](#), Neha Ojha)
- core: osd/PG: fix misused FORCE\_RECOVERY[BACKFILL] flags ([issue#27985](#), [pr#23904](#), xie xingguo)
- core: osd/PGLog.cc: check if complete\_to points to log.end() ([pr#23450](#), Neha Ojha)
- core: osd/PGLog: trim - avoid dereferencing invalid iter ([pr#23546](#), xie xingguo)
- core: osd/PG: remove unused functions ([pr#26155](#), Kefu Chai)
- core: osd/PG: reset PG on osd down->up; normalize query processing ([issue#24373](#), [pr#22456](#), Sage Weil)
- core: osd/PG: restrict async\_recovery\_targets to up osds ([pr#22664](#), Neha Ojha)
- core: osd/PG: unset history\_les\_bound if local-les is used ([pr#22524](#), Kefu Chai)
- core: osd/PG: write pg epoch when resurrecting pg after delete vs merge race ([issue#35923](#), [pr#24061](#), Sage Weil)
- core: osd/PrimaryLogPG: do not count failed read in delta\_stats ([pr#25687](#), Kefu Chai)

- core: osd/PrimaryLogPG: fix last\_peering\_reset checking on manifest flushing ([pr#26778](#), xie xingguo)
- core: osd/PrimaryLogPG: fix on\_local\_recover crash on stray clone ([pr#22396](#), Sage Weil)
- core: osd/PrimaryLogPG: fix potential pg-log overtrimming ([pr#23317](#), xie xingguo)
- core: osd/PrimaryLogPG: fix the extent length error of the sync read ([pr#25584](#), Xiaofei Cui)
- core: osd/PrimaryLogPG: fix try\_flush\_mark\_clean write contention case ([issue#24174](#), [pr#22084](#), Sage Weil)
- core: osd/PrimaryLogPG: optimize recover order ([pr#23587](#), xie xingguo)
- core: osd/PrimaryLogPG: update missing\_loc more carefully ([issue#35546](#), [pr#23895](#), xie xingguo)
- core: osd/ReplicatedBackend: remove useless assert ([pr#21243](#), Jianpeng Ma)
- core: osd/Session: fix invalid iterator dereference in Session::have\_backoff() ([issue#24486](#), [pr#22497](#), Sage Weil)
- core: osd: write “debug dump\_missing” output to stdout ([pr#21960](#), Коренберг Марк)
- core: os/kstore: support db statistic ([pr#21487](#), Yang Honggang)
- core: os/memstore: use ceph::mutex and friends ([pr#26026](#), Kefu Chai)
- core,performance: core: avoid unnecessary refcounting of OSDMap on OSD’s hot paths ([pr#24743](#), Radoslaw Zarzynski)
- core,performance: msg/async: avoid put message within write\_lock ([pr#20731](#), Haomai Wang)
- core,performance: os/bluestore: make osd shard-thread do oncommits ([pr#22739](#), Jianpeng Ma)
- core,performance: osd/filestore: Change default filestore\_merge\_threshold to -10 ([issue#24686](#), [pr#22761](#), Douglas Fuller)
- core,performance: osd/OSDMap: map pgs with smaller batchs in calc\_pg\_upmaps ([pr#23734](#), huangjun)
- core: PG: release reservations after backfill completes ([issue#23614](#), [pr#22255](#), Neha Ojha)
- core: pg stuck in backfill\_wait with plenty of disk space ([issue#38034](#), [pr#26375](#), xie xingguo, David Zafman)

- core,pybind: pybind/rados: new methods for manipulating self-managed snapshots ([pr#22579](#), Jason Dillaman)
- core: qa/suites/rados: minor fixes ([pr#22195](#), Neha Ojha)
- core: qa/suites/rados/thrash-erasure-code\*/thrashers/\*: less likely reserv rejection injection ([pr#24667](#), Sage Weil)
- core: qa/suites/rados/thrash-old-clients: only centos and 16.04 ([pr#22106](#), Sage Weil)
- core: qa/suites: set osd\_pg\_log\_dups\_tracked in cfuse\_workunit\_suites\_fsync.yaml ([pr#21909](#), Neha Ojha)
- core: qa/suites/upgrade/luminous-x: disable c-o-t import/export tests between versions ([issue#38294](#), [pr#27018](#), Sage Weil)
- core: qa/suites/upgrade/mimic-x/parallel: enable all classes ([pr#27011](#), Sage Weil)
- core: qa/workunits/mgr/test\_localpool.sh: use new config syntax ([pr#22496](#), Sage Weil)
- core: qa/workunits/rados/test\_health\_warnings: prevent out osds ([issue#37776](#), [pr#25732](#), Sage Weil)
- core: rados.pyx: make all exceptions accept keyword arguments ([issue#24033](#), [pr#21853](#), Rishabh Dave)
- core: rados: return legacy address in 'lock info' ([pr#26150](#), Jason Dillaman)
- core: scrub warning check incorrectly uses mon scrub interval ([issue#37264](#), [pr#25112](#), David Zafman)
- core: src: no 'dne' acronym in user cmd output ([pr#21094](#), Gu Zhongyan)
- core,tests: Minor cleanups in tests and log output ([issue#38631](#), [issue#38678](#), [pr#26899](#), David Zafman)
- core,tests: qa/overrides/short\_pg\_log.yaml: reduce osd\_{min,max}\_pg\_log\_entries ([issue#38025](#), [pr#26101](#), Neha Ojha)
- core,tests: qa/suites/rados/thrash: change crush\_tunables to jewel in rados\_api\_tests ([issue#38042](#), [pr#26122](#), Neha Ojha)
- core,tests: qa/suites/upgrade/luminous-x: a few fixes ([pr#22092](#), Sage Weil)
- core,tests: qa/tests: Set ansible-version: 2.5 ([issue#24926](#), [pr#23123](#), Yuri Weinstein)
- core,tests: Removal of snapshot with corrupt replica crashes osd ([issue#23875](#), [pr#22476](#), David Zafman)

- core,tests: test: Verify a log trim trims the dup\_index ([pr#26533](#), Brad Hubbard)
- core,tools: osdmaptool: fix wrong test\_map\_pgs\_dump\_all output ([pr#22280](#), huangjun)
- core,tools: rados: provide user with more meaningful error message ([pr#26275](#), Mykola Golub)
- core,tools: tools/rados: allow reuse object for write test ([pr#25128](#), Li Wang)
- core: vstart.sh: Support SPDK in Ceph development deployment ([pr#22975](#), tone.zhang)
- crimson: add MonClient ([pr#23849](#), Kefu Chai)
- crimson: cache osdmap using LRU cache ([pr#26254](#), Kefu Chai, Jianpeng Ma)
- crimson/common: apply config changes also on shard.0 ([pr#23631](#), Yingxin)
- crimson/connection: misc changes ([pr#23044](#), Kefu Chai)
- crimson: crimson/mon: remove timeout support from mon::Client::authenticate() ([pr#24660](#), Kefu Chai)
- crimson/mon: move mon::Connection into .cc ([pr#24619](#), Kefu Chai)
- crimson/net: concurrent dispatch for SocketMessenger ([pr#24090](#), Casey Bodley)
- crimson/net: encapsulate protocol implementations with states ([pr#25176](#), Yingxin, Kefu Chai)
- crimson/net: encapsulate protocol implementations with states (remaining part) ([pr#25207](#), Yingxin)
- crimson/net: fix addresses during banner exchange ([pr#25580](#), Yingxin)
- crimson/net: fix compile errors in test\_alien\_echo.cc ([pr#24629](#), Yingxin)
- crimson/net: fix crimson msgr error leaks to caller ([pr#25716](#), Yingxin)
- crimson/net: fix misc issues for segment-fault and test-failures ([pr#25939](#), Yingxin Cheng, Kefu Chai)
- crimson/net: Fix racing for promise on\_message ([pr#24097](#), Yingxin)
- crimson/net: fix unittest\_seastar\_messenger errors ([pr#23539](#), Yingxin)
- crimson/net: implement accepting/connecting states ([pr#24608](#), Yingxin)
- crimson/net: miscellaneous fixes to seastar-msgr ([pr#23816](#), Yingxin, Casey Bodley)
- crimson/net: misc fixes and features for crimson-messenger tests ([pr#26221](#),

Yingxin Cheng)

- crimson/net: seastar-msgr refactoring ([pr#24576](#), Yingxin)
- crimson/net: s/repeat/keep\_doing/ ([pr#23898](#), Kefu Chai)
- crimson/osd: add heartbeat support ([pr#26222](#), Kefu Chai)
- crimson/osd: add more heartbeat peers ([pr#26255](#), Kefu Chai)
- crimson/osd: correct the order of parameters passed to OSD::\_preboot() ([pr#26774](#), chunmei Liu)
- crimson/osd: crimson osd driver ([pr#25304](#), Radoslaw Zarzynski, Kefu Chai)
- crimson/osd: remove "force\_new" from ms\_get\_authorizer() ([pr#26054](#), Kefu Chai)
- crimson/osd: send known addresses at boot ([pr#26452](#), Kefu Chai)
- crimson: persist/load osdmap to/from store ([pr#26090](#), Kefu Chai)
- crimson: port messenger to seastar ([pr#22491](#), Kefu Chai, Casey Bodley)
- crimson/thread: add thread pool ([pr#22565](#), Kefu Chai)
- crimson/thread: pin thread pool to given CPU ([pr#22776](#), Kefu Chai)
- crush/CrushWrapper: silence compiler warning ([pr#25336](#), Li Wang)
- crush: fix device\_class\_clone for unpopulated/empty weight-sets ([issue#23386](#), [pr#22127](#), Sage Weil)
- crush: fix memory leak ([pr#25959](#), xie xingguo)
- crush: fix upmap overkill ([issue#37968](#), [pr#26179](#), xie xingguo)
- dashboard/mgr: Save button doesn't prevent saving an invalid form ([issue#36426](#), [pr#24577](#), Patrick Nawracay)
- dashboard: Return float if rate not available ([pr#22313](#), Boris Ranto)
- doc: add Ceph Manager Dashboard to top-level TOC ([pr#26390](#), Nathan Cutler)
- doc: add ceph-volume inventory sections ([pr#25092](#), Jan Fajerski)
- doc: add documentation for iostat ([pr#22034](#), Mohamad Gebai)
- doc: added demo document changes section ([pr#24791](#), James McClune)
- doc: added rbd default features ([pr#24720](#), Gaurav Sitlani)
- doc: added some Civetweb configuration options ([pr#24073](#), Anton Oks)
- doc: Added some hints on how to further accelerate builds with ccache ([pr#25394](#),

Lenz Grimmer)

- doc: add instructions about using “serve-doc” to preview built document ([pr#24471](#), Kefu Chai)
- doc: add mds state transition diagram ([issue#22989](#), [pr#22996](#), Patrick Donnelly)
- doc: Add mention of ceph osd pool stats ([pr#25575](#), Thore Kruess)
- doc: add missing 12.2.11 release note ([pr#26596](#), Nathan Cutler)
- doc: add note about LVM volumes to ceph-deploy quick start ([pr#23879](#), David Wahler)
- doc: add release notes for 12.2.11 luminous ([pr#26228](#), Abhishek Lekshmanan)
- doc: add spacing to subcommand references ([pr#24669](#), James McClune)
- doc: add “-timeout” option to rbd-nbd ([pr#24302](#), Stefan Kooman)
- doc/bluestore: fix minor typos in compression section ([pr#22874](#), David Disseldorp)
- doc: broken link on troubleshooting-mon page ([pr#25312](#), James McClune)
- doc: bump up sphinx and pyyaml versions ([pr#26044](#), Kefu Chai)
- doc: ceph-deploy would not support -cluster option anymore ([pr#26471](#), Tatsuya Naganawa)
- doc: ceph: describe application subcommand in ceph man page ([pr#20645](#), Rishabh Dave)
- doc: ceph-iscsi-api ports should not be public facing ([pr#24248](#), Jason Dillaman)
- doc: ceph-volume describe better the options for migrating away from ceph-disk ([issue#24036](#), [pr#21890](#), Alfredo Deza)
- doc: ceph-volume dmcrypt and activate -all documentation updates ([issue#24031](#), [pr#22062](#), Alfredo Deza)
- doc: ceph-volume: expand on why ceph-disk was replaced ([pr#23194](#), Alfredo Deza)
- doc: ceph-volume: lvm batch documentation and man page updates ([issue#24970](#), [pr#23443](#), Alfredo Deza)
- doc: ceph-volume: update batch documentation to explain filestore strategies ([issue#34309](#), [pr#23785](#), Alfredo Deza)
- doc: ceph-volume: zfs, the initial first submit ([pr#23674](#), Willem Jan Withagen)
- doc: cleaned up troubleshooting OSDs documentation ([pr#23519](#), James McClune)

- doc: Clean up field names in ServiceDescription and add a service field ([pr#26006](#), Jeff Layton)
- doc: cleanup: prune Argonaut-specific verbiage ([pr#22899](#), Nathan Cutler)
- doc: cleanup rendering syntax ([pr#22389](#), Mahati Chamarthry)
- doc: Clean up the snapshot consistency note ([pr#25655](#), Greg Farnum)
- doc: common,mon: add implicit #include headers ([pr#23930](#), Kefu Chai)
- doc: common/options: add description of osd objectstore backends ([issue#24147](#), [pr#22040](#), Alfredo Deza)
- doc: corrected options of iscsiadm command ([pr#26395](#), ZhuJieWen)
- doc: correct rbytes description ([pr#24966](#), Xiang Dai)
- doc: describe RBD QoS settings ([pr#25202](#), Mykola Golub)
- doc: doc/bluestore: data doesn't use two partitions (ceph-disk era) ([pr#22604](#), Alfredo Deza)
- doc: doc/cephfs: fixup add/remove mds docs ([pr#23836](#), liu wei)
- doc: doc/cephfs: remove lingering "experimental" note about multimds ([pr#22852](#), John Spray)
- doc: doc/dashboard: don't advise mgr\_initial\_modules ([pr#22808](#), John Spray)
- doc: doc/dashboard: fix formatting on Grafana instructions-2 ([pr#22706](#), Jos Collin)
- doc: doc/dashboard: fix formatting on Grafana instructions ([pr#22657](#), John Spray)
- doc: doc/dev/cephx\_protocol: fix couple errors ([pr#23750](#), Kefu Chai)
- doc: doc/dev/index: update rados lead ([pr#24160](#), Josh Durgin)
- doc: doc/dev/msgr2.rst: update of the banner and authentication phases ([pr#20094](#), Ricardo Dias)
- doc: doc/dev/seastore.rst: initial draft notes ([pr#21381](#), Sage Weil)
- doc: doc/dev: Updated component leads table ([pr#24238](#), Lenz Grimmer)
- doc: doc: fix the links in releases/schedule.rst ([pr#22364](#), Kefu Chai)
- doc: doc/man: mention import and export commands in rados manpage ([issue#4640](#), [pr#23186](#), Nathan Cutler)
- doc: doc: Mention PURGED\_SNAPDIRS and RECOVERY\_DELETE in Mimic release notes ([pr#22711](#), Florian Haas)

- doc: doc/mgr/dashboard: fix typo in mgr ssl setup ([pr#24790](#), Mehdi Abaakouk)
- doc: doc/mgr: mention how to clear config setting ([pr#22157](#), John Spray)
- doc: doc/mgr: note need for module.py file in plugins ([pr#22622](#), John Spray)
- doc: doc/mgr/orchestrator: Add Architecture Image ([pr#26331](#), Sebastian Wagner, Kefu Chai)
- doc: doc/mgr/orchestrator: add wal to blink lights ([pr#25634](#), Sebastian Wagner)
- doc: doc/mgr/prometheus: readd section about custom instance labels ([pr#25182](#), Jan Fajerski)
- doc: doc/orchestrator: Aligned Documentation with specification ([pr#25893](#), Sebastian Wagner)
- doc: doc/orchestrator: Integrate CLI specification into the documentation ([pr#25119](#), Sebastian Wagner)
- doc: doc: purge subcommand link broken ([pr#24785](#), James McClune)
- doc: doc/rados: Add bluestore memory autotuning docs ([pr#25069](#), Mark Nelson)
- doc: doc/rados/configuration: add osd scrub {begin,end} week day ([pr#25924](#), Neha Ojha)
- doc: doc/rados/configuration/msgr2: some documentation about msgr2 ([pr#26867](#), Sage Weil)
- doc: doc/rados/configuration: refresh osdmap section ([pr#26120](#), Ilya Dryomov)
- doc: doc/rados: correct osd path in troubleshooting-mon.rst ([pr#24964](#), songweibin)
- doc: doc/rados: fixed hit set type link ([pr#23833](#), James McClune)
- doc: doc/radosgw/s3.rst: Adding AWS S3 Storage Class as Not Supported ([pr#19571](#), Katie Holly)
- doc: doc/rados/operations: add balancer.rst to TOC ([pr#23684](#), Kefu Chai)
- doc: doc/rados/operations: add clay to erasure-code-profile ([pr#26902](#), Kefu Chai)
- doc: doc/rados/operations/crush-map-edits: fix 'take' syntax ([pr#24868](#), Remy Zandwijk, Sage Weil)
- doc: doc/rados/operations/pg-states: fix PG state names, part 2 ([pr#23165](#), Nathan Cutler)
- doc: doc/rados/operations/pg-states: fix PG state names ([pr#21520](#), Jan Fajerski)

- doc: doc/rados update invalid bash on bluestore migration ([issue#34317](#), [pr#23801](#), Alfredo Deza)
- doc: doc/rbd: corrected OpenStack Cinder permissions for Glance pool ([pr#22443](#), Jason Dillaman)
- doc: doc/rbd: explicitly state that mirroring requires connectivity to clusters ([pr#24433](#), Jason Dillaman)
- doc: doc/rbd/iscsi-target-cli: Update auth command ([pr#26788](#), Ricardo Marques)
- doc: doc/rbd/iscsi-target-cli: Update disk separator ([pr#26669](#), Ricardo Marques)
- doc: doc/release/luminous: v12.2.6 and v12.2.7 release notes ([pr#23057](#), Abhishek Lekshmanan, Sage Weil)
- doc: doc/releases: Add luminous releases 12.2.9 and 10 ([pr#25361](#), Brad Hubbard)
- doc: doc/releases: Add Mimic release 13.2.2 ([pr#24509](#), Brad Hubbard)
- doc: doc/releases: Mark Jewel EOL ([pr#23698](#), Brad Hubbard)
- doc: doc/releases: Mark Mimic first release as June ([pr#24099](#), Brad Hubbard)
- doc: doc/releases/mimic.rst: make note of 13.2.2 upgrade bug ([pr#24979](#), Neha Ojha)
- doc: doc/releases/mimic: tweak RBD major features ([pr#22011](#), Jason Dillaman)
- doc: doc/releases/mimic: Updated dashboard description ([pr#22016](#), Lenz Grimmer)
- doc: doc/releases/mimic: upgrade steps ([pr#21987](#), Sage Weil)
- doc: doc/releases/nautilus: dashboard package notes ([pr#26815](#), Kefu Chai)
- doc: doc/releases/schedule: Add Luminous 12.2.8 ([pr#23972](#), Brad Hubbard)
- doc: doc/releases/schedule: add mimic column ([pr#22006](#), Sage Weil)
- doc: doc/releases: Update releases to August '18 ([pr#23360](#), Brad Hubbard)
- doc: doc/rgw: document placement targets and storage classes ([issue#24508](#), [issue#38008](#), [pr#26997](#), Casey Bodley)
- doc: docs: add Clay code plugin documentation ([pr#24422](#), Myna)
- doc: docs: Fixed swift client authentication fail ([pr#23729](#), Dai Dang Van)
- doc: docs: radosgw: ldap-auth: fixed option name 'rgw\_ldap\_searchfilter' ([issue#23081](#), [pr#20526](#), Konstantin Shalygin)
- doc: doc/start: fix kube-helm.rst typo: docuiment -> document ([pr#23423](#), Zhou Peng)

- doc: doc/SubmittingPatches.rst: use Google style guide for doc patches ([pr#22190](#), Nathan Cutler)
- doc: Document correction ([pr#23926](#), Gangbiao Liu)
- doc: Document mappings of S3 Operations to ACL grants ([pr#26827](#), Adam C. Emerson)
- doc: document sizing for block.db ([pr#23210](#), Alfredo Deza)
- doc: document vstart options ([pr#22467](#), Mao Zhongyi)
- doc: doc/user-management: Remove obsolete reset caps command ([issue#37663](#), [pr#25550](#), Brad Hubbard)
- doc: edit on github ([pr#24452](#), Neha Ojha, Noah Watkins)
- doc: erasure-code-clay fixes typos ([pr#24653](#), Myna)
- doc: erasure-code-jerasure: removed default section of crush-device-class ([pr#21279](#), Junyoung Sung)
- doc: examples/librados: Remove not needed else clauses ([pr#24939](#), Marcos Paulo de Souza)
- doc: explain 'firstn v indep' in the CRUSH docs ([pr#24255](#), Greg Farnum)
- doc: Fix a couple typos and improve diagram formatting ([pr#23496](#), Bryan Stillwell)
- doc: fix a typo in doc/mgr/telegraf.rst ([pr#22267](#), Enming Zhang)
- doc: fix cephfs spelling errors ([pr#23763](#), Chen Zhenghua)
- doc: fix/cleanup freebsd osd disk creation ([pr#23600](#), Willem Jan Withagen)
- doc: Fix Create a Cluster url in Running Multiple Clusters ([issue#37764](#), [pr#25705](#), Jos Collin)
- doc: Fix EC k=3 m=2 profile overhead calculation example ([pr#20581](#), Charles Alva)
- doc: fixed broken urls ([pr#23564](#), James McClune)
- doc: fixed grammar in restore rbd image section ([pr#22944](#), James McClune)
- doc: fixed links in Pools section ([pr#23431](#), James McClune)
- doc: fixed minor typo in Debian packages section ([pr#22878](#), James McClune)
- doc: fixed restful mgr module SSL configuration commands ([pr#21864](#), Lenz Grimmer)
- doc: Fixed spelling errors in configuration section ([pr#23719](#), Bryan Stillwell)
- doc: Fixed syntax in iscsi initiator windows doc ([pr#25467](#), Michel Raabe)

- doc: Fixed the paragraph and boxes ([pr#25094](#), Scoots Hamilton)
- doc: Fixed the wrong numbers in mgr/dashboard.rst ([pr#22658](#), Jos Collin)
- doc: fixed typo in add-or-rm-mons.rst ([pr#26250](#), James McClune)
- doc: fixed typo in cephfs snapshots ([pr#23764](#), Kai Wagner)
- doc: fixed typo in CRUSH map docs ([pr#25953](#), James McClune)
- doc: fixed typo in man page ([pr#24792](#), James McClune)
- doc: Fix incorrect mention of ‘osd\_deep\_mon\_scrub\_interval’ ([pr#26522](#), Ashish Singh)
- doc: Fix iSCSI docs URL ([pr#26296](#), Ricardo Marques)
- doc: fix iscsi target name when configuring target ([pr#21906](#), Venky Shankar)
- doc: fix long description error for rgw\_period\_root\_pool ([pr#23814](#), yuliyang)
- doc: fix some it's -> its typos ([pr#22802](#), Brad Fitzpatrick)
- doc: Fix some typos ([pr#25060](#), mooncake)
- doc: Fix Spelling Error In File “ceph.rst” ([pr#23917](#), Gangbiao Liu)
- doc: Fix Spelling Error In File dynamicresharding.rst ([pr#24175](#), xiaomanh)
- doc: Fix Spelling Error of Rados Deployment/Operations ([pr#23746](#), Li Bingyang)
- doc: Fix Spelling Error of Radosgw ([pr#23948](#), Li Bingyang)
- doc: Fix Spelling Error of Radosgw ([pr#24000](#), Li Bingyang)
- doc: Fix Spelling Error of Radosgw ([pr#24021](#), Li Bingyang)
- doc: Fix Spelling Error of Rados Operations ([pr#23891](#), Li Bingyang)
- doc: Fix Spelling Error of Rados Operations ([pr#23900](#), Li Bingyang)
- doc: Fix Spelling Error of Rados Operations ([pr#23903](#), Li Bingyang)
- doc: fix spelling errors in rbd doc ([pr#23765](#), Chen Zhenghua)
- doc: fix spelling errors of cephfs ([pr#23745](#), Chen Zhenghua)
- doc: fix the broken urls ([issue#25185](#), [pr#23310](#), Jos Collin)
- doc: fix the formatting of HTTP Frontends documentation ([pr#25723](#), James McClune)
- doc: fix typo and format issues in quick start documentation ([pr#23705](#), Chen Zhenghua)

- doc: fix typo in add-or-rm-mons ([pr#25661](#), Jos Collin)
- doc: Fix typo in ceph-fuse(8) ([pr#22214](#), Jos Collin)
- doc: fix typo in erasure coding example ([pr#25737](#), Arthur Liu)
- doc: Fix typos in Developer Guide ([pr#24067](#), Li Bingyang)
- doc: fix typos in doc/releases ([pr#24186](#), Li Bingyang)
- doc: \*/: fix typos in docs, messages, logs, comments ([pr#24139](#), Kefu Chai)
- doc: Fix Typos of Developer Guide ([pr#24094](#), Li Bingyang)
- doc: fix typos ([pr#22174](#), Mao Zhongyi)
- doc: .githubmap, .mailmap, .organizationmap: update contributors ([pr#24756](#), Tiago Melo)
- doc: githubmap, organizationmap: cleanup and add/update contributors/affiliation ([pr#22734](#), Tatjana Dehler)
- doc: give pool name if default pool rbd is not created ([pr#24750](#), Changcheng Liu)
- doc: Improve docs osd\_recovery\_priority, osd\_recovery\_op\_priority and related ([pr#26705](#), David Zafman)
- doc: Improve OpenStack integration and multitenancy docs for radosgw ([issue#36765](#), [pr#25056](#), Florian Haas)
- doc: install build-doc deps without git clone ([pr#24416](#), Noah Watkins)
- doc: Luminous v12.2.10 release notes ([pr#25034](#), Nathan Cutler)
- doc: Luminous v12.2.9 release notes ([pr#24779](#), Nathan Cutler)
- doc: make it easier to reach the old dev doc TOC ([pr#23253](#), Nathan Cutler)
- doc: mention CVEs in luminous v12.2.11 release notes ([pr#26312](#), Nathan Cutler, Abhishek Lekshmanan)
- doc: mgr/dashboard: Add documentation about supported browsers ([issue#27207](#), [pr#23712](#), Tiago Melo)
- doc: mgr/dashboard: Added missing tooltip to settings icon ([pr#23935](#), Lenz Grimmer)
- doc: mgr/dashboard: Add hints to resolve unit test failures ([pr#23627](#), Stephan Müller)
- doc: mgr/dashboard: Cleaner notifications ([pr#23315](#), Stephan Müller)
- doc: mgr/dashboard: Cleanup of summary refresh test ([pr#25504](#), Stephan Müller)

- doc: mgr/dashboard: Document custom RESTController endpoints ([pr#25322](#), Stephan Müller)
- doc: mgr/dashboard: Fixed documentation link on RGW page ([pr#24612](#), Tina Kallio)
- doc: mgr/dashboard: Fix some setup steps in HACKING.rst ([pr#24788](#), Ranjitha G)
- doc: mgr/dashboard: Improve prettier scripts and documentation ([pr#22994](#), Tiago Melo)
- doc: mgr/dashboard/qa: add missing dashboard suites ([pr#25084](#), Tatjana Dehler)
- doc: mgr/dashboard: updated SSO documentation ([pr#25943](#), Alfonso Martínez)
- doc: mgr/dashboard: Update I18N documentation ([pr#25159](#), Tiago Melo)
- doc: mgr/orch: Fix remote\_host doc reference ([issue#38254](#), [pr#26360](#), Ernesto Puerta)
- doc/mgr/plugins.rst: explain more about the plugin command protocol ([pr#22629](#), Dan Mick)
- doc: mimic is stable! ([pr#22350](#), Abhishek Lekshmanan)
- doc: mimic rc1 release notes ([pr#20975](#), Abhishek Lekshmanan)
- doc: Multiple spelling fixes ([pr#23514](#), Bryan Stillwell)
- doc: numbered eviction situations ([pr#24618](#), Scoots Hamilton)
- doc: osdmaptool/cleanup: Completed osdmaptool's usage ([issue#3214](#), [pr#13925](#), Vedant Nanda)
- doc: osd/PrimaryLogPG: avoid dereferencing invalid complete\_to ([pr#23894](#), xie xingguo)
- doc: osd/PrimaryLogPG: rename list\_missing -> list\_unfound command ([pr#23723](#), xie xingguo)
- doc: PendingReleaseNotes: note newly added CLAY code ([pr#24491](#), Kefu Chai)
- doc: print pg peering in SVG instead of PNG ([pr#20366](#), Aleksei Gutikov)
- doc: Put command template into literal block ([pr#24999](#), Alexey Stupnikov)
- doc: qa/mgr/selftest: handle always-on module fall out ([issue#26994](#), [pr#23681](#), Noah Watkins)
- doc: qa: Task to emulate network delay and packet drop between two given h... ([pr#23602](#), Shilpa Jagannath)
- doc: qa/workunits/rbd: replace usage of 'rados rmpool' ([pr#23942](#), Mykola Golub)

- doc: release/mimic: correct the changelog to the latest version ([pr#22319](#), Abhishek Lekshmanan)
- doc: release notes for 12.2.8 luminous ([pr#23909](#), Abhishek Lekshmanan)
- doc: release notes for 13.2.2 mimic ([pr#24266](#), Abhishek Lekshmanan)
- doc: releases: mimic 13.2.1 release notes ([pr#23288](#), Abhishek Lekshmanan)
- doc: releases: release notes for v10.2.11 Jewel ([pr#22989](#), Abhishek Lekshmanan)
- doc: remove CZ mirror ([pr#21797](#), Tomáš Kukrál)
- doc: remove deprecated 'scrubq' from ceph(8) ([issue#35813](#), [pr#23959](#), Ruben Kerkhof)
- doc: remove documentation for installing google-perf-tools on Debian systems ([pr#22701](#), James McClune)
- doc: remove duplicate python packages ([pr#22203](#), Stefan Kooman)
- doc: Remove upstart files and references ([pr#23582](#), Brad Hubbard)
- doc: Remove value 'mon\_osd\_max\_split\_count' ([pr#26584](#), Kai Wagner)
- doc: replace rgw\_namespace\_expire\_secs with rgw\_nfs\_namespace\_expire\_secs ([pr#20794](#), chnmagnus)
- doc: rewrote the iscsi-target-cli installation ([pr#23190](#), Massimiliano Cuttini)
- doc: rgw: fix tagging support status ([issue#24164](#), [pr#22206](#), Abhishek Lekshmanan)
- doc: rgw: fix the default value of usage log setting ([issue#37856](#), [pr#25892](#), Abhishek Lekshmanan)
- doc: Rook/orchestrator doc fixes ([pr#23472](#), John Spray)
- doc: s/doc/ref for dashboard urls ([pr#22772](#), Jos Collin)
- doc: sort releases by date and version ([pr#25972](#), Noah Watkins)
- doc: Spelling fixes in BlueStore config reference ([pr#23715](#), Bryan Stillwell)
- doc: Spelling fixes in Network config reference ([pr#23727](#), libingyang)
- doc: SubmittingPatches: added inline markup to important references ([pr#25978](#), James McClune)
- docs: update rgw info for mimic ([pr#22305](#), Yehuda Sadeh)
- doc: test/crimson: do not use unit.cc as the driver of unittest\_seastar\_denc ([pr#23937](#), Kefu Chai)

- doc: test/fio: Added tips for compilation of fio with 'rados' engine ([pr#24199](#), Adam Kupczyk)
- doc: test/msgr: add missing #include ([pr#23947](#), Kefu Chai)
- doc: Tidy up description wording and spelling ([pr#22599](#), Anthony D'Atri)
- doc: tweak RBD iSCSI docs to point to merged tooling repo ([pr#24963](#), Jason Dillaman)
- doc: typo fixes, s/Requered/Required/ ([pr#26406](#), Drunkard Zhang)
- doc: update blkin changes ([pr#22317](#), Mahati Chamopathy)
- doc: Update cpp.rst to accommodate the new APIs in libs3 ([pr#22162](#), Zhanhao Liu)
- doc: Updated Ceph Dashboard documentation ([pr#26626](#), Lenz Grimmer)
- doc: updated Ceph documentation links ([pr#25797](#), James McClune)
- doc: updated cluster map reference link ([pr#24460](#), James McClune)
- doc: updated crush map tunables link ([pr#24462](#), James McClune)
- doc: Updated dashboard documentation (features, SSL config) ([pr#22059](#), Lenz Grimmer)
- doc: Updated feature list and overview in dashboard.rst ([pr#26143](#), Lenz Grimmer)
- doc: updated get-involved.rst for ceph-dashboard ([pr#22663](#), Jos Collin)
- doc: Updated Mgr Dashboard documentation ([pr#24030](#), Lenz Grimmer)
- doc: updated multisite documentation ([issue#26997](#), [pr#23660](#), James McClune)
- doc: updated reference link for creating new disk offerings in cloudstack ([pr#22250](#), James McClune)
- doc: updated reference link for log based PG ([pr#26611](#), James McClune)
- doc: updated rgw multitenancy link ([pr#25929](#), James McClune)
- doc: updated the overview and glossary for dashboard ([pr#22750](#), Jos Collin)
- doc: updated wording from federated to multisite ([pr#24670](#), James McClune)
- doc: Update mgr/zabbix plugin documentation with link to Zabbix template ([pr#24584](#), Wido den Hollander)
- doc: update the description for SPDK in bluestore-config-ref.rst ([pr#22365](#), tone-zhang)
- doc: use :command: for subcommands in ceph-bluestore-tool manpage ([issue#24800](#),

[pr#23114](#), Nathan Cutler)

- doc: use preferred commands for ceph config-key ([pr#26527](#), Changcheng Liu)
- doc: warn about how ‘rados put’ works in the manpage ([pr#25757](#), Greg Farnum)
- doc: Wip githubmap ([pr#25950](#), Greg Farnum)
- erasure-code,test: silence -Wunused-variable warnings ([pr#25200](#), Kefu Chai)
- example/librados: remove dependency on Boost system library ([issue#25054](#), [pr#23159](#), Nathan Cutler)
- githubmap: update contributors ([pr#22522](#), Kefu Chai)
- git: Ignore tags anywhere ([pr#26159](#), David Zafman)
- include/buffer.h: do not use ceph\_assert() unless \_\_CEPH\_\_ is defined ([pr#23803](#), Kefu Chai)
- install-deps.sh: Fixes for RHEL 7 ([pr#26393](#), Zack Cerza)
- kv/MemDB: add perfcounter ([pr#10305](#), Jianpeng Ma)
- librados: add a rados\_omap\_iter\_size function ([issue#26948](#), [pr#23593](#), Jeff Layton)
- librados: block MgrClient::start\_command until mgrmap ([pr#21811](#), John Spray, Kefu Chai)
- librados: fix admin/build-doc warning ([pr#25706](#), Jos Collin)
- librados: fix buffer overflow for aio\_exec python binding ([pr#21775](#), Aleksei Gutikov)
- librados: fix uninitialized timeout in wait\_for\_osdmap ([pr#24721](#), Casey Bodley)
- librados: Include memory for unique\_ptr definition ([issue#35833](#), [pr#23992](#), Brad Hubbard)
- librados: Reject the invalid pool create request at client side, rath... ([pr#21299](#), Yang Honggang)
- librados: return ENOENT if pool\_id invalid ([pr#21609](#), Li Wang)
- librados: split C++ and C APIs into different source files ([pr#24616](#), Kefu Chai)
- librados: use ceph::async::Completion for asio bindings ([pr#21920](#), Casey Bodley)
- librados: use steady clock for rados\_mon\_op\_timeout ([pr#20004](#), Mohamad Gebai)
- librbd: add missing shutdown states to managed lock helper ([issue#38387](#), [pr#26523](#), Jason Dillaman)

- librbd: add new configuration option to always move deleted items to the trash ([pr#24476](#), Jason Dillaman)
- librbd: add rbd image access/modified timestamps ([pr#21114](#), Julien Collet)
- librbd: add trash purge api calls ([pr#24427](#), Julien Collet, Theofilos Mouratidis, Jason Dillaman)
- librbd: always open first parent image if it exists for a snapshot ([pr#23733](#), Jason Dillaman)
- librbd: avoid aggregate-initializing any static\_visitor ([pr#26876](#), Willem Jan Withagen)
- librbd: blacklisted client might not notice it lost the lock ([issue#34534](#), [pr#23829](#), Jason Dillaman)
- librbd: block\_name\_prefix is not created randomly ([issue#24634](#), [pr#22675](#), hyun-ha)
- librbd: bypass pool validation if "rbd\_validate\_pool" is false ([pr#26878](#), Jason Dillaman)
- librbd: commit IO as safe when complete if writeback cache is disabled ([issue#23516](#), [pr#22342](#), Jason Dillaman)
- librbd: corrected usage of ImageState::open flag parameter ([pr#25428](#), Mykola Golub)
- librbd: deep\_copy: don't hide parent if zero overlap for snapshot ([issue#24545](#), [pr#22587](#), Mykola Golub)
- librbd: deep copy optionally support flattening cloned image ([issue#22787](#), [pr#21624](#), Mykola Golub)
- librbd: deep\_copy: resize head object map if needed ([issue#24399](#), [pr#22415](#), Mykola Golub)
- librbd: deep-copy should not write to objects that cannot exist ([issue#25000](#), [pr#23132](#), Jason Dillaman)
- librbd: disable image mirroring when moving to trash ([pr#25509](#), Mykola Golub)
- librbd: disallow trash restoring when image being migrated ([pr#25529](#), songweibin)
- librbd: don't do create+truncate for discards with copyup ([pr#26825](#), Ilya Dryomov)
- librbd: ensure compare-and-write doesn't skip compare after copyup ([issue#38383](#), [pr#26519](#), Ilya Dryomov)
- librbd: extend API to include parent/child namespaces and image ids ([issue#36650](#),

- pr#25194, Jason Dillaman)
- librbd: fix crash when opening nonexistent snapshot ([issue#24637](#), [pr#22676](#), Mykola Golub)
  - librbd: fixed assert when flattening clone with zero overlap ([issue#35702](#), [pr#24045](#), Jason Dillaman)
  - librbd: fix missing unblock\_writes if shrink is not allowed ([issue#36778](#), [pr#25055](#), runsi)
  - librbd: fix possible unnecessary latency when requeue request ([pr#23815](#), Song Shun)
  - librbd: fix potential live migration after commit issues due to not refreshed image header ([pr#23839](#), Mykola Golub)
  - librbd: fix were\_all\_throttled() to avoid incorrect ret-value ([issue#38504](#), [pr#26688](#), Dongsheng Yang)
  - librbd: flatten operation should use object map ([issue#23445](#), [pr#23941](#), Mykola Golub)
  - librbd: force 'invalid object map' flag on-disk update ([issue#24434](#), [pr#22444](#), Mykola Golub)
  - librbd: get\_parent API method should properly handle migrating image ([issue#37998](#), [pr#26337](#), Jason Dillaman)
  - librbd: handle aio failure in ManagedLock and PreReleaseRequest ([pr#20112](#), liyichao)
  - librbd: improve object map performance under high IOPS workloads ([issue#38538](#), [pr#26721](#), Jason Dillaman)
  - librbd: journaling unable request can not be sent to remote lock owner ([issue#26939](#), [pr#23649](#), Mykola Golub)
  - librbd: keep access/modified timestamp updates out of IO path ([issue#37745](#), [pr#25883](#), Jason Dillaman)
  - librbd: make it possible to migrate parent images ([pr#25945](#), Mykola Golub)
  - librbd: move mirror peer attribute handling from CLI to API ([pr#25096](#), Jason Dillaman)
  - librbd: namespace create/remove/list support ([pr#22608](#), Jason Dillaman)
  - librbd: object copy state machine might dereference a deleted object ([issue#36220](#), [pr#24293](#), Jason Dillaman)
  - librbd: object map improperly flagged as invalidated ([issue#24516](#), [pr#24105](#),

Jason Dillaman)

- librbd: optionally limit journal in-flight appends ([pr#22983](#), Mykola Golub)
- librbd: optionally support FUA (force unit access) on write requests ([issue#19366](#), [pr#22945](#), ningtao)
- librbd: pool and image level config overrides ([pr#23743](#), Mykola Golub)
- librbd: potential object map race with copyup state machine ([issue#24516](#), [pr#24253](#), Jason Dillaman)
- librbd: potential race on image create request complete ([issue#24910](#), [pr#23639](#), Mykola Golub)
- librbd: prevent the use of internal feature bits from external users ([issue#24165](#), [pr#22072](#), Jason Dillaman)
- librbd: prevent use of namespaces on pre-nautilus OSDs ([pr#23823](#), Jason Dillaman)
- librbd: properly filter out trashed non-user images on purge ([pr#26079](#), Mykola Golub)
- librbd: properly handle potential object map failures ([issue#36074](#), [pr#24179](#), Jason Dillaman)
- librbd: race condition possible when validating RBD pool ([issue#38500](#), [pr#26683](#), Jason Dillaman)
- librbd: reduce the TokenBucket fill cycle and support bursting io configuration ([pr#24214](#), Shiyang Ruan)
- librbd: remove template declaration of a non-template function ([pr#23790](#), Shiyang Ruan)
- librbd: reset snaps in rbd\_snap\_list() ([issue#37508](#), [pr#25379](#), Kefu Chai)
- librbd: restart io if migration parent gone ([issue#36710](#), [pr#25175](#), Mykola Golub)
- librbd: send\_copyup() fixes and cleanups ([pr#26483](#), Ilya Dryomov)
- librbd: simplify config override handling ([pr#24450](#), Jason Dillaman)
- librbd: skip small, unaligned discard extents by default ([issue#38146](#), [pr#26432](#), Jason Dillaman)
- librbd: support bps throttle and throttle read and write separately ([pr#21635](#), Dongsheng Yang)
- librbd: support migrating images with minimal downtime ([issue#18430](#), [issue#24439](#), [issue#26874](#), [issue#23659](#), [pr#15831](#), Patrick Donnelly, Sage Weil, Alfredo Deza, Kefu Chai, Patrick Nawracay, Pavani Rajula, Mykola Golub, Casey Bodley, Yingxin,

Jason Dillaman)

- librbd: support v2 cloning across namespaces ([pr#23662](#), Jason Dillaman)
- librbd: use object map when doing snap rollback ([pr#23110](#), songweibin)
- librbd: utilize the journal disabled policy when removing images ([issue#23512](#), [pr#22327](#), Jason Dillaman)
- librbd: validate data pool for self-managed snapshot support ([pr#22737](#), Mykola Golub)
- librbd: workaround an ICE of GCC ([issue#37719](#), [pr#25733](#), Kefu Chai)
- log: avoid heap allocations for most log entries ([pr#23721](#), Patrick Donnelly)
- lvm: when osd creation fails log the exception ([issue#24456](#), [pr#22627](#), Andrew Schoen)
- mailmap,organization: Update sangfor affiliation ([pr#25225](#), Zengran Zhang)
- mds: add reference when setting Connection::priv to existing session ([pr#22384](#), "Yan, Zheng")
- mds: fix leak of MDSCacheObject::waiting ([issue#24289](#), [pr#22307](#), "Yan, Zheng")
- mds: fix some memory leak ([issue#24289](#), [pr#22240](#), "Yan, Zheng")
- mds,messages: silence -Wclass-memaccess warnings ([pr#21845](#), Kefu Chai)
- mds: properly journal root inode's snaprealm ([issue#24343](#), [pr#22320](#), "Yan, Zheng")
- mds: remove obsolete comments ([pr#25549](#), Patrick Donnelly)
- mds: reply session reject for open request from blacklisted client ([pr#21941](#), Yan, Zheng, "Yan, Zheng")
- mgr: Add ability to trigger a cluster/audit log message from Python ([pr#24239](#), Volker Theile)
- mgr: Add HandleCommandResult namedtuple ([pr#25261](#), Sebastian Wagner)
- mgr: add limit param to osd perf query ([pr#25151](#), Mykola Golub)
- mgr: add per pool force-recovery/backfill commands ([issue#38456](#), [pr#26560](#), xie xingguo)
- mgr: add per pool scrub commands ([pr#26532](#), xie xingguo)
- mgr: Allow modules to get/set other module options ([pr#25651](#), Volker Theile)
- mgr: Allow rook to scale the mon count ([pr#26405](#), Jeff Layton)

- mgr: always on modules v2 ([pr#23970](#), Noah Watkins)
- mgr/ansible: Add/remove hosts ([pr#26241](#), Juan Miguel Olmo Martínez)
- mgr/ansible: Replace Ansible playbook used to retrieve storage devices data ([pr#26023](#), Juan Miguel Olmo Martínez)
- mgr/ansible: Replace deprecated <get\_config> calls ([pr#25964](#), Juan Miguel Olmo Martínez)
- mgr: Centralize PG\_STATES to MgrModule ([pr#22594](#), Wido den Hollander)
- mgr: ceph-mgr: hold lock while accessing the request list and submitting request ([pr#25048](#), Jerry Lee)
- mgr: change 'bytes' dynamic perf counters to COUNTER type ([pr#25908](#), Mykola Golub)
- mgr: create always on class of modules ([pr#23106](#), Noah Watkins)
- mgr: create shell OSD performance query class ([pr#24117](#), Mykola Golub)
- mgr/dashboard: About modal proposed changes ([issue#35693](#), [pr#25376](#), Kanika Murarka)
- mgr/dashboard: Add ability to list, set and unset cluster-wide OSD flags to the backend ([issue#24056](#), [pr#21998](#), Patrick Nawracay)
- mgr/dashboard: Add a 'clear filter' button to configuration page ([issue#36173](#), [pr#25712](#), familyuu)
- mgr/dashboard: add a script to run an API request on a rook cluster ([pr#25991](#), Jeff Layton)
- mgr/dashboard: Add a unit test form helper class ([pr#24633](#), Stephan Müller)
- mgr/dashboard: Add backend support for changing dashboard configuration settings via the REST API ([pr#22457](#), Patrick Nawracay)
- mgr/dashboard: Add breadcrumbs component ([issue#24781](#), [pr#23414](#), Tiago Melo)
- mgr/dashboard: add columns to Pools table ([pr#25791](#), Alfonso Martínez)
- mgr/dashboard: Add decorator to skip parameter encoding ([issue#26856](#), [pr#23419](#), Tiago Melo)
- mgr/dashboard: Add description to menu items on mobile navigation ([pr#26198](#), Sebastian Krah)
- mgr/dashboard: added command to tox.ini ([pr#26073](#), Alfonso Martínez)
- mgr/dashboard: added 'env\_build' to 'npm run e2e' ([pr#26165](#), Alfonso Martínez)

- mgr/dashboard: Added new validators ([pr#22526](#), Stephan Müller)
- mgr/dashboard: Add error handling on the frontend ([pr#21820](#), Tiago Melo)
- mgr/dashboard: add Feature Toggles ([issue#37530](#), [pr#26102](#), Ernesto Puerta)
- mgr/dashboard: Add Filesystems list component ([pr#21913](#), Tiago Melo)
- mgr/dashboard: Add filtered rows number in table footer ([pr#22504](#), Tiago Melo)
- mgr/dashboard: Add gap between panel footer buttons ([pr#23796](#), Volker Theile)
- mgr/dashboard: Add guideline how to brand the UI and update the color scheme ([pr#25988](#), Sebastian Krah)
- mgr/dashboard: Add help menu entry ([pr#22303](#), Ricardo Marques)
- mgr/dashboard: Add i18n support ([pr#24803](#), Sebastian Krah, Tiago Melo)
- mgr/dashboard: Add implicit wait in e2e tests ([pr#26384](#), Tiago Melo)
- mgr/dashboard: Add info to Pools table ([pr#25489](#), Alfonso Martínez)
- mgr/dashboard: Add iSCSI discovery authentication UI ([pr#26320](#), Tiago Melo)
- mgr/dashboard: Add iSCSI Target Edit UI ([issue#38014](#), [pr#26367](#), Tiago Melo)
- mgr/dashboard: Add left padding to helper icon ([pr#24631](#), Stephan Müller)
- mgr/dashboard: Add missing frontend I18N ([issue#36719](#), [pr#25654](#), Tiago Melo)
- mgr/dashboard: Add missing test requirement "werkzeug" ([pr#24628](#), Stephan Müller)
- mgr/dashboard: Add NFS status endpoint ([issue#38399](#), [pr#26539](#), Tiago Melo)
- mgr/dashboard: Add 'no-unused-variable' rule to tslint ([pr#22328](#), Tiago Melo)
- mgr/dashboard: Add permission validation to the "Purge Trash" button ([issue#36272](#), [pr#24370](#), Tiago Melo)
- mgr/dashboard: Add pool cache tiering details tab ([issue#25158](#), [pr#25602](#), familyuu)
- mgr/dashboard: Add Pool update endpoint ([pr#21881](#), Sebastian Wagner, Stephan Müller)
- mgr/dashboard: Add Prettier formatter to the frontend ([pr#21819](#), Tiago Melo)
- mgr/dashboard: add profiles to set cluster's rebuild performance ([pr#24968](#), Tatjana Dehler)
- mgr/dashboard: add pytest plugin: faulthandler ([pr#25053](#), Alfonso Martínez)

- mgr/dashboard: Add REST API for role management ([pr#23322](#), Ricardo Marques)
- mgr/dashboard: Add scrub action to the OSDs table ([pr#22122](#), Tiago Melo)
- mgr/dashboard: Adds custom timepicker for grafana iframes ([pr#25583](#), Kanika Murarka)
- mgr/dashboard: Adds ECP management to the frontend ([pr#24627](#), Stephan Müller)
- mgr/dashboard: Add shared Confirmation Modal ([pr#22601](#), Tiago Melo)
- mgr/dashboard: add supported flag information to config options documentation ([pr#22760](#), Tatjana Dehler)
- mgr/dashboard: Add support for iSCSI's multi backstores (UI) ([pr#26575](#), Tiago Melo)
- mgr/dashboard: Add support for managing individual OSD settings/characteristics in the frontend ([issue#36487](#), [issue#36444](#), [issue#35448](#), [issue#36188](#), [issue#35811](#), [issue#35816](#), [issue#36086](#), [pr#24606](#), Patrick Nawracay)
- mgr/dashboard: Add support for managing individual OSD settings in the backend ([issue#24270](#), [pr#23491](#), Patrick Nawracay)
- mgr/dashboard: Add support for managing RBD QoS ([issue#37572](#), [issue#38004](#), [issue#37570](#), [issue#37936](#), [issue#37574](#), [issue#36191](#), [issue#37845](#), [issue#37569](#), [pr#25233](#), Patrick Nawracay)
- mgr/dashboard: Add support for RBD Trash ([issue#24272](#), [pr#23351](#), Tiago Melo)
- mgr/dashboard: Add support for URI encode ([issue#24621](#), [pr#22672](#), Tiago Melo)
- mgr/dashboard: Add table actions component ([pr#23779](#), Stephan Müller)
- mgr/dashboard: Add table of contents to HACKING.rst ([pr#25812](#), Sebastian Krah)
- mgr/dashboard: Add token authentication to Grafana proxy ([pr#22459](#), Patrick Nawracay)
- mgr/dashboard: Add TSLint rule “no-unused-variable” ([pr#24699](#), Alfonso Martínez)
- mgr/dashboard: Add UI for Cluster-wide OSD Flags configuration ([pr#22461](#), Tiago Melo)
- mgr/dashboard: Add UI for disabling ACL authentication ([issue#38218](#), [pr#26388](#), Tiago Melo)
- mgr/dashboard: Add UI to configure the telemetry mgr plugin ([pr#25989](#), Volker Theile)
- mgr/dashboard: Add unique validator ([pr#23802](#), Volker Theile)

- mgr/dashboard: Allow “/” in pool name ([issue#38302](#), [pr#26408](#), Tiago Melo)
- mgr/dashboard: Allow insecure HTTPS in run-backend-api-request ([pr#21882](#), Sebastian Wagner)
- mgr/dashboard: Allow renaming an existing Pool ([issue#36560](#), [pr#25107](#), guodan1)
- mgr/dashboard: Audit REST API calls ([pr#24475](#), Volker Theile)
- mgr/dashboard: Auto-create a name for RBD image snapshots ([pr#23735](#), Volker Theile)
- mgr/dashboard: avoid blank content in Read/Write Card ([pr#25563](#), Alfonso Martínez)
- mgr/dashboard: awssauth: fix python3 string decode problem ([pr#21794](#), Ricardo Dias)
- mgr/dashboard: Can't handle user editing when tenants are specified ([pr#24757](#), Volker Theile)
- mgr/dashboard: Catch LookupError when checking the RGW status ([pr#24028](#), Volker Theile)
- mgr/dashboard: CdFormGroup ([pr#22644](#), Stephan Müller)
- mgr/dashboard: Ceph dashboard user management from the UI ([pr#22758](#), Ricardo Marques)

- mgr/dashboard: Change 'Client Recovery' title ([pr#26883](#), Ernesto Puerta)
- mgr/dashboard: Changed background color of Masthead to brand gray ([issue#35690](#), [pr#25628](#), Neha Gupta)
- mgr/dashboard: Changed default value of decimal point to 1 ([pr#22386](#), Tiago Melo)
- mgr/dashboard: Change icon color in notifications ([pr#26586](#), Volker Theile)
- mgr/dashboard: Check content-type before decode json response ([pr#24350](#), Ricardo Marques)
- mgr/dashboard: check for existence of Grafana dashboard ([issue#36356](#), [pr#25154](#), Kanika Murarka)
- mgr/dashboard: Cleanup of OSD list methods ([pr#24823](#), Stephan Müller)
- mgr/dashboard: Cleanup of the cluster and audit log ([pr#26188](#), Sebastian Krah)
- mgr/dashboard: Cleanup ([pr#24831](#), Patrick Nawracay)
- mgr/dashboard: Clean up pylint's disable:no-else-return ([pr#26509](#), Patrick Nawracay)
- mgr/dashboard: Cleanup Python code ([pr#26743](#), Volker Theile)
- mgr/dashboard: Cleanup RGW config checks ([pr#22669](#), Volker Theile)
- mgr/dashboard: Close modal dialogs on login screen ([pr#23328](#), Volker Theile)
- mgr/dashboard: code cleanup ([pr#25502](#), Alfonso Martínez)
- mgr/dashboard: Color variables for color codes ([issue#24575](#), [pr#22695](#), Kanika Murarka)
- mgr/dashboard config options add ([issue#34528](#), [issue#24996](#), [issue#24455](#), [issue#36173](#), [pr#23230](#), Tatjana Dehler)
- mgr/dashboard: Config options integration (read-only) depends on #22422 ([pr#21460](#), Tatjana Dehler)
- mgr/dashboard: config options table cleanup ([issue#34533](#), [pr#24523](#), Tatjana Dehler)
- mgr/dashboard: config option type names update ([issue#37843](#), [pr#25876](#), Tatjana Dehler)
- mgr/dashboard: configs textarea disallow horizontal resize ([issue#36452](#), [pr#24614](#), Tatjana Dehler)
- mgr/dashboard: Configure all mgr modules in UI ([pr#26116](#), Volker Theile)

- mgr/dashboard: Confirmation modal doesn't close ([pr#24544](#), Volker Theile)
- mgr/dashboard: Confusing tilted time stamps in the CephFS performance graph ([pr#25909](#), Volker Theile)
- mgr/dashboard: consider config option default values ([issue#37683](#), [pr#25616](#), Tatjana Dehler)
- mgr/dashboard: controller infrastructure refactor and new features ([pr#22210](#), Patrick Nawracay, Ricardo Dias)
- mgr/dashboard: Correct permission decorator ([pr#26135](#), Tina Kallio)
- mgr/dashboard: CRUSH map viewer ([issue#35684](#), [pr#24766](#), familyuu)
- mgr/dashboard: CRUSH map viewer RFE ([issue#37794](#), [pr#26162](#), familyuu)
- mgr/dashboard: Dashboard info cards refactoring ([pr#22902](#), Alfonso Martínez)
- mgr/dashboard: Datatable error panel blinking on page loading ([pr#23316](#), Volker Theile)
- mgr/dashboard: Deletion dialog falsely executes deletion when pressing 'Cancel' ([pr#22003](#), Volker Theile)
- mgr/dashboard: Disable package-lock.json creation ([pr#22061](#), Tiago Melo)
- mgr/dashboard: Disable RBD actions during task execution ([pr#23445](#), Ricardo Marques)
- mgr/dashboard: disallow editing read-only config options (part 2) ([pr#26450](#), Tatjana Dehler)
- mgr/dashboard: disallow editing read-only config options ([pr#26297](#), Tatjana Dehler)
- mgr/dashboard: Display logged in user ([issue#24822](#), [pr#24213](#), guodan1, guodan)
- mgr/dashboard: Display notification if RGW is not configured ([pr#21785](#), Volker Theile)
- mgr/dashboard: Display RGW user/bucket quota max size in human readable form ([pr#23842](#), Volker Theile)
- mgr/dashboard: Do not fetch pool list on RBD edit ([pr#22404](#), Ricardo Marques)
- mgr/dashboard: Do not require cert for http ([issue#36069](#), [pr#24103](#), Boris Ranto)
- mgr/dashboard: Drop iSCSI gateway name parameter ([pr#26984](#), Ricardo Marques)
- mgr/dashboard: enable coverage for API tests ([pr#26851](#), Alfonso Martínez)

- mgr/dashboard: Escape regex pattern in DeletionModalComponent ([issue#24902](#), [pr#23420](#), Tiago Melo)
- mgr/dashboard: Exception.message doesn't exist on Python 3 ([pr#24349](#), Ricardo Marques)
- mgr/dashboard: Extract/Refactor Task merge ([pr#23555](#), Stephan Müller, Tiago Melo)
- mgr/dashboard: Filter out tasks depending on permissions ([pr#25426](#), Tina Kallio)
- mgr/dashboard: Fix /api/grafana/validation ([pr#25997](#), Zack Cerza)
- mgr/dashboard: Fix bug in user form when changing password ([pr#23939](#), Volker Theile)
- mgr/dashboard: Fix cherrypy static content URL prefix config ([pr#23183](#), Ricardo Marques)
- mgr/dashboard: Fix duplicate error messages ([pr#23287](#), Stephan Müller)
- mgr/dashboard: Fix duplicate tasks ([pr#24930](#), Tiago Melo)
- mgr/dashboard: Fix e2e script ([pr#22903](#), Tiago Melo)
- mgr/dashboard: Fixed performance details context for host list row selection ([issue#37854](#), [pr#26020](#), Neha Gupta)
- mgr/dashboard: Fixed typos in environment.build.js ([pr#26650](#), Lenz Grimmer)
- mgr/dashboard: Fix error when clicking on newly created OSD ([issue#36245](#), [pr#24369](#), Patrick Nawracay)
- mgr/dashboard: Fixes documentation link- to open in new tab ([pr#22237](#), a2batic)
- mgr/dashboard: Fixes Grafana 500 error ([issue#37809](#), [pr#25830](#), Kanika Murarka)
- mgr/dashboard: Fix failing QA test: test\_safe\_to\_destroy ([issue#37290](#), [pr#25149](#), Patrick Nawracay)
- mgr/dashboard: Fix flaky QA tests ([pr#24024](#), Patrick Nawracay)
- mgr/dashboard: Fix Forbidden Error with some roles ([issue#37293](#), [pr#25141](#), Ernesto Puerta)
- mgr/dashboard: fix for 'Cluster >> Hosts' page ([pr#24974](#), Alfonso Martínez)
- mgr/dashboard: Fix formatter service unit test ([pr#22323](#), Tiago Melo)
- mgr/dashboard: fix for using '::' on hosts without ipv6 ([pr#26635](#), Noah Watkins)
- mgr/dashboard: Fix growing table in firefox ([issue#26999](#), [pr#23711](#), Tiago Melo)
- mgr/dashboard: Fix HttpClient Module imports in unit tests ([pr#24679](#), Tiago Melo)

- mgr/dashboard: Fix iSCSI mutual password input type ([pr#26854](#), Ricardo Marques)
- mgr/dashboard: Fix iSCSI service unit tests ([pr#26319](#), Tiago Melo)
- mgr/dashboard: Fix issues in controllers/docs ([pr#26738](#), Volker Theile)
- mgr/dashboard: Fix Jest conflict with coverage files ([pr#22155](#), Tiago Melo)
- mgr/dashboard: Fix layout issues in UI ([issue#24525](#), [pr#22597](#), Volker Theile)
- mgr/dashboard: Fix links to external documentation ([pr#24829](#), Patrick Nawracay)
- mgr/dashboard: fix lint error caused by codelyzer update ([pr#22693](#), Tiago Melo)
- mgr/dashboard: fix lint error ([pr#22417](#), Tiago Melo)
- mgr/dashboard: Fix long running RBD cloning / copying message ([pr#24641](#), Ricardo Marques)
- mgr/dashboard: Fix missing failed restore notification ([issue#36513](#), [pr#24664](#), Tiago Melo)
- mgr/dashboard: Fix modified files only (frontend) ([pr#25346](#), Patrick Nawracay)
- mgr/dashboard: Fix moment.js deprecation warning ([pr#21981](#), Tiago Melo)
- mgr/dashboard: Fix more layout issues in UI ([pr#22600](#), Volker Theile)
- mgr/dashboard: Fix navbar focused color ([pr#25769](#), Volker Theile)
- mgr/dashboard: Fix notifications in user list and form ([pr#23797](#), Volker Theile)
- mgr/dashboard: Fix OSD down error display ([issue#24530](#), [pr#23754](#), Patrick Nawracay)
- mgr/dashboard: Fix pool usage not displaying on filesystem page ([pr#22453](#), Tiago Melo)
- mgr/dashboard: Fix problem with ErasureCodeProfileService ([pr#24694](#), Tiago Melo)
- mgr/dashboard: Fix Python3 issue ([pr#24617](#), Patrick Nawracay)
- mgr/dashboard: fix query parameters in task annotated endpoints ([issue#25096](#), [pr#23229](#), Ricardo Dias)
- mgr/dashboard: Fix RBD actions disable ([pr#24637](#), Ricardo Marques)
- mgr/dashboard: Fix RBD features style ([pr#22759](#), Ricardo Marques)
- mgr/dashboard: Fix RBD object size dropdown options ([pr#22830](#), Ricardo Marques)
- mgr/dashboard: Fix RBD task metadata ([pr#22088](#), Tiago Melo)

- mgr/dashboard: Fix redirect to login page on session lost ([pr#23388](#), Ricardo Marques)
- mgr/dashboard: fix reference to oA ([pr#24343](#), Joao Eduardo Luis)
- mgr/dashboard: Fix regression on rbd form component ([issue#24757](#), [pr#22829](#), Tiago Melo)
- mgr/dashboard: Fix reloading of pool listing ([pr#26182](#), Patrick Nawracay)
- mgr/dashboard: Fix renaming of pools ([pr#25423](#), Patrick Nawracay)
- mgr/dashboard: Fix search in Source column of RBD configuration list ([issue#37569](#), [pr#26765](#), Patrick Nawracay)
- mgr/dashboard: fix skipped backend API tests ([pr#26172](#), Alfonso Martínez)
- mgr/dashboard: Fix some datatable CSS issues ([pr#22216](#), Volker Theile)
- mgr/dashboard: Fix spaces around status labels on OSD list ([pr#24607](#), Patrick Nawracay)
- mgr/dashboard: Fix summary refresh call stack ([pr#25984](#), Tiago Melo)
- mgr/dashboard: Fix test\_full\_health test ([issue#37872](#), [pr#25913](#), Tatjana Dehler)
- mgr/dashboard: Fix test\_remove\_not\_expired\_trash qa test ([issue#37354](#), [pr#25221](#), Tiago Melo)
- mgr/dashboard: fix: toast notifications hiding utility menu ([pr#26429](#), Alfonso Martínez)
- mgr/dashboard: fix: tox not detecting deps changes ([pr#26409](#), Alfonso Martínez)
- mgr/dashboard: Fix ts error on iSCSI page ([pr#24715](#), Ricardo Marques)
- mgr/dashboard: Fix typo in NoOrchesrtatorConfiguredException class name ([pr#26334](#), Volker Theile)
- mgr/dashboard: Fix typo in pools management ([pr#26323](#), Lenz Grimmer)
- mgr/dashboard: Fix typo ([pr#23363](#), Volker Theile)
- mgr/dashboard: Fix unit tests cli warnings ([pr#21933](#), Tiago Melo)
- mgr/dashboard: Format small numbers correctly ([issue#24081](#), [pr#21980](#), Stephan Müller)
- mgr/dashboard: Get user ID via RGW Admin Ops API ([pr#22416](#), Volker Theile)
- mgr/dashboard: Grafana dashboard updates and additions ([pr#24314](#), Paul Cuzner)
- mgr/dashboard: Grafana graphs integration with dashboard ([pr#23666](#), Kanika

Murarka)

- mgr/dashboard: Grafana proxy backend ([pr#21644](#), Patrick Nawracay)
- mgr/dashboard: Group buttons together into one menu on OSD page ([issue#37380](#), [pr#26189](#), Tatjana Dehler)
- mgr/dashboard: Handle class objects as regular objects in KV-table ([pr#24632](#), Stephan Müller)
- mgr/dashboard: Handle errors during deletion ([pr#22002](#), Volker Theile)
- mgr/dashboard: Hide empty fields and render all objects in KV-table ([pr#25894](#), Stephan Müller)
- mgr/dashboard: Hide progress bar in case of an error ([pr#22419](#), Volker Theile)
- mgr/dashboard: Implement OSD purge ([issue#35811](#), [pr#26242](#), Patrick Nawracay)
- mgr/dashboard: Improve CRUSH map viewer ([pr#24934](#), Volker Theile)
- mgr/dashboard: Improved support for generating OpenAPI Spec documentation ([issue#24763](#), [pr#26227](#), Tina Kallio)
- mgr/dashboard: Improve error message handling ([pr#24322](#), Volker Theile)
- mgr/dashboard: Improve error panel ([pr#21851](#), Volker Theile)
- mgr/dashboard: Improve exception handling in /api/rgw/status ([pr#25836](#), Volker Theile)
- mgr/dashboard: Improve exception handling ([issue#23823](#), [pr#21066](#), Sebastian Wagner)
- mgr/dashboard: Improve HACKING.rst ([pr#22281](#), Patrick Nawracay)
- mgr/dashboard: Improve 'no pool' message on rbd form ([pr#22150](#), Ricardo Marques)
- mgr/dashboard: Improve RBD form ([issue#38303](#), [pr#26433](#), Tiago Melo)
- mgr/dashboard: Improve RGW address parser ([pr#25870](#), Volker Theile)
- mgr/dashboard: Improve RgwUser controller ([pr#25300](#), Volker Theile)
- mgr/dashboard: Improves documentation for Grafana Setting ([issue#36371](#), [pr#24511](#), Kanika Murarka)
- mgr/dashboard: Improve str\_to\_bool ([pr#22757](#), Volker Theile)
- mgr/dashboard: Improve SummaryService and TaskwrapperService ([pr#22906](#), Tiago Melo)
- mgr/dashboard: Improve table pagination style ([pr#22065](#), Ricardo Marques)

- mgr/dashboard: Introduce pipe to convert bool to text ([pr#26507](#), Volker Theile)
- mgr/dashboard: iscsi: adds CLI command to enable/disable API SSL verification ([pr#26891](#), Ricardo Dias)
- mgr/dashboard: iSCSI - Adds support for pool/image names with dots ([pr#26503](#), Ricardo Marques)
- mgr/dashboard: iSCSI - Add support for disabling ACL authentication (backend) ([pr#26382](#), Ricardo Marques)
- mgr/dashboard: iSCSI discovery authentication API ([pr#26115](#), Ricardo Marques)
- mgr/dashboard: iSCSI - Infrastructure for multiple backstores (backend) ([pr#26506](#), Ricardo Marques)
- mgr/dashboard: iSCSI management API ([pr#25638](#), Ricardo Marques, Ricardo Dias)
- mgr/dashboard: iSCSI management UI ([pr#25995](#), Ricardo Marques, Tiago Melo)
- mgr/dashboard: iSCSI - Support iSCSI passwords with '/' ([pr#26790](#), Ricardo Marques)
- mgr/dashboard: JWT authentication ([pr#22833](#), Ricardo Dias)
- mgr/dashboard: Landing Page: chart improvements ([pr#24810](#), Alfonso Martínez)
- mgr/dashboard: Landing Page: info visibility ([pr#24513](#), Alfonso Martínez)
- mgr/dashboard: Log frontend errors + @UiController ([pr#22285](#), Ricardo Marques)
- mgr/dashboard: Login failure should return HTTP 400 ([pr#22403](#), Ricardo Marques)
- mgr/dashboard: 'Logs' links permission in Landing Page ([pr#25231](#), Alfonso Martínez)
- mgr/dashboard: Make deletion dialog more touch device friendly ([pr#23897](#), Volker Theile)
- mgr/dashboard: Map dev 'releases' to master ([pr#24763](#), Zack Cerza)
- mgr/dashboard: Module dashboard.services.ganesha has several lint issues ([pr#26378](#), Volker Theile)
- mgr/dashboard: More configs for table updateSelectionOnRefresh ([pr#24015](#), Ricardo Marques)
- mgr/dashboard: Move Cluster/Audit logs from front page to dedicated Logs page ([pr#23834](#), Diksha Godbole)
- mgr/dashboard: Move unit-test-helper into the new testing folder ([pr#22857](#), Tiago Melo)

- mgr/dashboard: Navbar dropdown button does not respond for mobile browsers ([pr#21967](#), Volker Theile)
- mgr/dashboard: New Landing Page: Milestone 2 ([pr#24326](#), Alfonso Martínez)
- mgr/dashboard: New Landing Page ([pr#23568](#), Alfonso Martínez)
- mgr/dashboard: nfs-ganesha: controller API documentation ([pr#26716](#), Ricardo Dias)
- mgr/dashboard: NFS management UI ([pr#26085](#), Tiago Melo)
- mgr/dashboard: ng serve bind to 0.0.0.0 ([pr#22058](#), Ricardo Marques)
- mgr/dashboard: no side-effects on failed user creation ([pr#24200](#), Joao Eduardo Luis)
- mgr/dashboard: Notification queue ([pr#25325](#), Stephan Müller)
- mgr/dashboard: npm run e2e:dev ([pr#25136](#), Stephan Müller)
- mgr/dashboard: Performance counter progress bar keeps infinitely looping ([pr#24448](#), Volker Theile)
- mgr/dashboard: permanent pie chart slice hiding ([pr#25276](#), Alfonso Martínez)
- mgr/dashboard: PGs will update as expected ([pr#26589](#), Stephan Müller)
- mgr/dashboard: Pool management ([pr#21614](#), Stephan Müller)
- mgr/dashboard: pool stats not returned by default ([pr#25635](#), Alfonso Martínez)
- mgr/dashboard: Possible fix for some dashboard timing issues ([issue#36107](#), [pr#24219](#), Patrick Nawracay)
- mgr/dashboard: Prettify package.json ([pr#22401](#), Ricardo Marques)
- mgr/dashboard: Prettify RGW JS code ([pr#22278](#), Volker Theile)
- mgr/dashboard: Prevent API call on every keystroke ([pr#23391](#), Volker Theile)
- mgr/dashboard: Print a blank space between value and unit ([pr#22387](#), Volker Theile)
- mgr/dashboard: Progress bar does not stop in TableKeyValueComponent ([pr#24016](#), Volker Theile)
- mgr/dashboard: Prometheus integration ([pr#25309](#), Stephan Müller)
- mgr/dashboard: Provide all four 'mandatory' OSD flags ([issue#37857](#), [pr#25905](#), Tatjana Dehler)
- mgr/dashboard/qa: Fix ECP creation test ([pr#25120](#), Stephan Müller)

- mgr/dashboard/qa: Fix various vstart\_runner.py issues ([issue#36581](#), [pr#24767](#), Volker Theile)
- mgr/dashboard: Redirect /block to /block/rbd ([pr#24722](#), Zack Cerza)
- mgr/dashboard: Reduce Jest logs in CI ([pr#24764](#), Tiago Melo)
- mgr/dashboard: Refactor autofocus directive ([pr#23910](#), Volker Theile)
- mgr/dashboard: Refactoring of DeletionModalComponent ([pr#24005](#), Patrick Nawracay)
- mgr/dashboard: Refactor perf counters ([pr#21673](#), Volker Theile)
- mgr/dashboard: Refactor RGW backend ([pr#21784](#), Volker Theile)
- mgr/dashboard: Refactor role management ([pr#23960](#), Volker Theile)
- mgr/dashboard: Relocate empty pipe ([pr#26588](#), Volker Theile)
- mgr/dashboard: Removed unnecessary fake services from unit tests ([pr#22473](#), Stephan Müller)
- mgr/dashboard: Remove fieldsets when using CdTable ([pr#23730](#), Tiago Melo)
- mgr/dashboard: Remove \_filterValue from CdFormGroup ([issue#26861](#), [pr#24719](#), Stephan Müller)
- mgr/dashboard: Remove husky package ([pr#21971](#), Tiago Melo)
- mgr/dashboard: Remove karma packages ([pr#23181](#), Tiago Melo)
- mgr/dashboard: Remove param when calling notificationService.show ([pr#26447](#), Volker Theile)
- mgr/dashboard: Remove top-right actions text and add “About” page ([pr#22762](#), Ricardo Marques)
- mgr/dashboard: Remove unused code ([pr#25439](#), Patrick Nawracay)
- mgr/dashboard: Remove useless code ([pr#23911](#), Volker Theile)
- mgr/dashboard: Remove useless observable unsubscriptions ([pr#21928](#), Ricardo Marques)
- mgr/dashboard: replace configuration html table with cd-table ([pr#21643](#), Tatjana Dehler)
- mgr/dashboard: Replaced “Pool” with “Pools” in navigation bar ([pr#22715](#), Lenz Grimmer)
- mgr/dashboard: Replace RGW proxy controller ([issue#24436](#), [pr#22470](#), Volker Theile)

- mgr/dashboard: Reset settings to their default values ([pr#22298](#), Patrick Nawracay)
- mgr/dashboard: Resolve TestBed performance issue ([pr#21783](#), Stephan Müller)
- mgr/dashboard: rest: add support for query params ([pr#22318](#), Ricardo Dias)
- mgr/dashboard: RestClient can't handle ProtocolError exceptions ([pr#23347](#), Volker Theile)
- mgr/dashboard: restcontroller: minor improvements and bug fixes ([pr#22528](#), Ricardo Dias)
- mgr/dashboard: RGW is not working if an URL prefix is defined ([pr#23200](#), Volker Theile)
- mgr/dashboard: RGW proxy can't handle self-signed SSL certificates ([pr#22735](#), Volker Theile)
- mgr/dashboard: role based authentication/authorization system ([issue#23796](#), [pr#22283](#), Ricardo Marques, Ricardo Dias)
- mgr/dashboard: Role management from the UI ([pr#23409](#), Ricardo Marques)
- mgr/dashboard: Search broken for entries with null values ([issue#38583](#), [pr#26766](#), Patrick Nawracay)
- mgr/dashboard: set errno via the parent class ([pr#21945](#), Kefu Chai, Ricardo Dias)
- mgr/dashboard: Set MODULE\_OPTIONS types and defaults ([pr#26386](#), Volker Theile)
- mgr/dashboard: Set timeout in RestClient calls ([pr#23224](#), Volker Theile)
- mgr/dashboard: Settings service ([pr#25327](#), Stephan Müller)
- mgr/dashboard: Show/Hide Grafana tabs according to user role ([issue#36655](#), [pr#24851](#), Kanika Murarka)
- mgr/dashboard: Show pool dropdown for block-mgr ([issue#37295](#), [pr#25144](#), Ernesto Puerta)
- mgr/dashboard: Show success notification in RGW forms ([pr#26482](#), Volker Theile)
- mgr/dashboard: Simplification of PoolForm method ([pr#24892](#), Patrick Nawracay)
- mgr/dashboard: Simplify OSD disabled action test ([pr#24824](#), Stephan Müller)
- mgr/dashboard: special casing for minikube in run-backend-rook-api-request.sh ([pr#26600](#), Jeff Layton)
- mgr/dashboard: SSO - SAML 2.0 support ([pr#24489](#), Ricardo Marques, Ricardo Dias)

- mgr/dashboard: SSO - UserDoesNotExist page ([pr#26058](#), Alfonso Martínez)
- mgr/dashboard: Stacktrace is optional on 'js-error' endpoint ([pr#22402](#), Ricardo Marques)
- mgr/dashboard: Status info cards' improvements ([pr#25155](#), Alfonso Martínez)
- mgr/dashboard: Store user table configurations ([pr#20822](#), Stephan Müller)
- mgr/dashboard: Stringify object[] in KV-table ([pr#22422](#), Stephan Müller)
- mgr/dashboard: Swagger-UI based Dashboard REST API page ([issue#23898](#), [pr#22282](#), Ricardo Dias)
- mgr/dashboard: Sync column style with the rest of the UI ([pr#26407](#), Volker Theile)
- mgr/dashboard: tasks.mgr.dashboard.test\_osd.OsdTest failures ([pr#24947](#), Volker Theile)
- mgr/dashboard: Task wrapper service ([pr#22014](#), Stephan Müller)
- mgr/dashboard: The RGW backend doesn't handle IPv6 properly ([pr#24222](#), Volker Theile)
- mgr/dashboard: typescript cleanup ([pr#26338](#), Alfonso Martínez)
- mgr/dashboard: Unit Tests cleanup ([pr#24591](#), Tiago Melo)
- mgr/dashboard: Update Angular packages ([pr#23706](#), Tiago Melo)
- mgr/dashboard: Update Angular to version 6 ([pr#22082](#), Tiago Melo)
- mgr/dashboard: Update bootstrap to v3.4.1 ([pr#26410](#), Tiago Melo)
- mgr/dashboard: Updated colors in PG Status chart ([pr#26203](#), Alfonso Martínez)
- mgr/dashboard: updated health API test ([pr#25813](#), Alfonso Martínez)
- mgr/dashboard: Updated image on 404 page ([pr#23820](#), Lenz Grimmer)
- mgr/dashboard: Update frontend packages ([pr#23466](#), Tiago Melo)
- mgr/dashboard: Update I18N translation ([pr#26649](#), Tiago Melo)
- mgr/dashboard: Update npm packages ([pr#24681](#), Tiago Melo)
- mgr/dashboard: Update npm packages ([pr#25656](#), Tiago Melo)
- mgr/dashboard: Update npm packages ([pr#26437](#), Tiago Melo)
- mgr/dashboard: Update npm packages ([pr#26647](#), Tiago Melo)

- mgr/dashboard: update python dependency ([pr#24928](#), Alfonso Martínez)
- mgr/dashboard: Update RxJS to version 6 ([pr#21826](#), Tiago Melo)
- mgr/dashboard: upgraded python dev dependencies ([pr#26007](#), Alfonso Martínez)
- mgr/dashboard: Upgrade Swimlane's data-table ([pr#21880](#), Volker Theile)
- mgr/dashboard: Use HTTPS in dev proxy configuration and HACKING.rst ([pr#21777](#), Volker Theile)
- mgr/dashboard: Use human readable units on the sparkline graphs ([issue#25075](#), [pr#23446](#), Tiago Melo)
- mgr/dashboard: User password should be optional ([pr#24128](#), Ricardo Marques)
- mgr/dashboard: Validate the OSD recovery priority form input values ([issue#37436](#), [pr#25472](#), Tatjana Dehler)
- mgr/dashboard: Validation for duplicate RGW user email ([issue#37369](#), [pr#25334](#), Kanika Murarka)
- mgr: define option defaults for MgrStandbyModule as well ([pr#25734](#), Kefu Chai)
- mgr: devicehealth: dont error on dict iteritems ([pr#22827](#), Abhishek Lekshmanan)
- mgr: Diskprediction cloud activate when config changes ([pr#25165](#), Rick Chen)
- mgr: don't write to output if EOPNOTSUPP ([issue#37444](#), [pr#25317](#), Kefu Chai)
- mgr: enable inter-module calls ([pr#22951](#), John Spray)
- mgr: Expose avgcount to the python modules ([pr#22010](#), Boris Ranto)
- mgr: expose avg data for long running avgs ([pr#22420](#), Boris Ranto)
- mgr: expose ec profiles through manager ([pr#23010](#), Noah Watkins)
- mgr: Extend batch to accept explicit device lists ([issue#37502](#), [issue#37086](#), [issue#37590](#), [pr#25542](#), Jan Fajerski)
- mgr: fix beacon interruption caused by deadlock ([pr#23482](#), Yan Jun)
- mgr: fix crash due to multiple sessions from daemons with same name ([pr#25534](#), Mykola Golub)
- mgr: fix permissions on balancer execute ([issue#25345](#), [pr#23387](#), John Spray)
- mgr: Fix rook spec and have service\_describe provide rados\_config\_location field for nfs services ([pr#25970](#), Jeff Layton)
- mgr: fix typo in variable name and cleanups ([pr#22069](#), Kefu Chai)

- mgr: fixup pgs show in unknown state ([issue#25103](#), [pr#23622](#), huanwen ren)
- mgr: Ignore daemon if no metadata was returned ([pr#22794](#), Wido den Hollander)
- mgr: Ignore `__pycache__` and wheelhouse dirs ([pr#26481](#), Volker Theile)
- mgr: Improve ActivePyModules::get\_typed\_config implementation ([pr#26149](#), Volker Theile)
- mgr: improve docs for MgrModule methods ([pr#22792](#), John Spray)
- mgr: improvements for dynamic osd perf counters ([pr#25488](#), Mykola Golub)
- mgr: Include daemon details in SLOW\_OPS output ([issue#23205](#), [pr#21750](#), Brad Hubbard)
- mgr: #include <vector> for clang ([pr#22756](#), Willem Jan Withagen)
- mgr: keep status, balancer always on ([pr#23558](#), Sage Weil)
- mgr: make module error message more descriptive ([pr#25537](#), Joao Eduardo Luis)
- mgr: mgr/ansible: Ansible orchestrator module ([pr#24445](#), Juan Miguel Olmo Martínez)
- mgr: mgr/ansible: Create/Remove OSDs ([pr#25497](#), Juan Miguel Olmo Martínez)
- mgr: mgr/ansible: Python 3 fix ([pr#25645](#), Sebastian Wagner)
- mgr: mgr/balancer: add min/max fields for weekday and be compatible with C ([pr#26505](#), xie xingguo)
- mgr: mgr/balancer: auto balance a list of pools ([pr#25940](#), xie xingguo)
- mgr: mgr/balancer: blame if upmap won't actually work ([pr#25941](#), xie xingguo)
- mgr: mgr/balancer: deepcopy best plan - otherwise we get latest ([issue#27000](#), [pr#23682](#), Stefan Priebe)
- mgr: mgr/balancer: restrict automatic balancing to specific weekdays ([pr#26440](#), xie xingguo)
- mgr: mgr/balancer: skip auto-balancing for pools with pending pg-merge ([pr#25626](#), xie xingguo)
- mgr: mgrc: enable disabling stats via `mgr_stats_threshold` ([issue#25197](#), [pr#23352](#), John Spray)
- mgr: mgr/crash: add hour granularity crash summary ([pr#23121](#), Noah Watkins)
- mgr: mgr/crash: add process name to crash metadata ([pr#25244](#), Mykola Golub)
- mgr: mgr/crash: fix python3 invalid syntax problems ([pr#23800](#), Ricardo Dias)

- mgr: mgr/DaemonServer: add js-output for "ceph osd safe-to-destroy" ([pr#24799](#), xie xingguo)
- mgr: mgr/DaemonServer: log pgmap usage to cluster log ([pr#26105](#), Neha Ojha)
- mgr: mgr/dashboard: Add option to disable SSL ([pr#22593](#), Wido den Hollander)
- mgr: mgr/dashboard: disable backend tests coverage ([pr#24193](#), Alfonso Martínez)
- mgr: mgr/dashboard: Fix dashboard shutdown/restart ([pr#22159](#), Boris Ranto)
- mgr: mgr/dashboard: Listen on port 8443 by default and not 8080 ([pr#22409](#), Wido den Hollander)
- mgr: mgr/dashboard: use the orchestrator\_cli backend setting ([pr#26325](#), Jeff Layton)
- mgr: mgr/deepsea: always use 'password' parameter for salt-api auth ([pr#26904](#), Tim Serong)
- mgr: mgr/deepsea: check for inflight completions when starting event reader, cleanup logging and comments ([pr#25391](#), Tim Serong)
- mgr: mgr/deepsea: DeepSea orchestrator module ([pr#24610](#), Tim Serong)
- mgr: mgr/devicehealth: clean up error handling ([pr#23205](#), John Spray)
- mgr: mgr/devicehealth: fix is\_valid\_daemon\_name typo error ([pr#24822](#), Lan Liu)
- mgr: mgr/diskprediction\_cloud: fix divide by zero when total\_size is 0 ([pr#26045](#), Rick Chen)
- mgr: mgr/diskprediction\_cloud: Remove needless library in the requirements file ([issue#37533](#), [pr#25433](#), Rick Chen)
- mgr: mgr/influx: Use Queue to store points which need to be written ([pr#23464](#), Wido den Hollander)
- mgr: mgr/insights: insights reporting module ([pr#23497](#), Noah Watkins)
- mgr: mgr/mgr\_module.py: fix doc for set\_store/set\_store\_json ([pr#22654](#), Dan Mick)
- mgr: mgr/orchestrator: Add RGW service support ([pr#23702](#), Rubab-Syed)
- mgr: mgr/orchestrator: Add service\_action method ([pr#25649](#), Tim Serong)
- mgr: mgr/orchestrator: Add support for "ceph orchestrator service ls" ([pr#24863](#), Jeff Layton)
- mgr: mgr/orchestrator: Improve debuggability ([pr#24147](#), Sebastian Wagner)
- mgr: mgr/orchestrator: Improve docstrings, add type hinting ([pr#25669](#), Sebastian

Wagner)

- mgr: mgr/orchestrator: Simplify Orchestrator wait implementation ([pr#25401](#), Juan Miguel Olmo Martinez)
- mgr: mgr/orchestrator: use result property in Completion classes ([pr#24672](#), Tim Serong)
- mgr: mgr/progress: improve+test OSD out handling ([pr#23146](#), John Spray)
- mgr: mgr/progress: introduce the progress module ([pr#22993](#), John Spray)
- mgr: mgr/prometheus: Add recovery metrics ([pr#26880](#), Paul Cuzner)
- mgr: mgr/prometheus: get osd\_objectstore once instead twice ([pr#26558](#), Konstantin Shalygin)
- mgr: mgr/restful: Fix deep-scrub typo ([issue#36720](#), [pr#24841](#), Boris Ranto)
- mgr: mgr/restful: fix py got exception when get osd info ([pr#21138](#), zouaiguo)
- mgr: mgr/restful: updated string formatting to str.format() ([pr#26210](#), James McClune)
- mgr: mgr/rook: fix API version and object types for recent rook changes ([pr#25452](#), Jeff Layton)
- mgr: mgr/rook: Fix Rook cluster name detection ([pr#24560](#), Sebastian Wagner)
- mgr: mgr/rook: update for v1beta1 API ([pr#23570](#), John Spray)
- mgr: mgr/status: Add standby-replay MDS ceph version ([pr#23624](#), Zhi Zhang)
- mgr: mgr/status: output to stdout, not stderr ([issue#24175](#), [pr#22089](#), John Spray)
- mgr: mgr/telegraf: Send more PG status information to Telegraf ([pr#22436](#), Wido den Hollander)
- mgr: mgr/telegraf: Telegraf module for Ceph Mgr ([pr#21782](#), Wido den Hollander)
- mgr: mgr/telegraf: Use Python generator and catch OSError ([pr#22418](#), Wido den Hollander)
- mgr: mgr/telemetry: Add Ceph Telemetry module to send reports back to project ([pr#21982](#), Wido den Hollander)
- mgr: mgr/telemetry: Check if boolean is False or not present ([pr#22223](#), Wido den Hollander)
- mgr: mgr/telemetry: Fix various issues ([pr#25770](#), Volker Theile)
- mgr: mgr/volumes: fix orchestrator remove operation ([pr#25339](#), Jeff Layton)

- mgr: mgr/zabbix: drop "total\_objects" field ([pr#26052](#), Kefu Chai)
- mgr: mgr/zabbix: Send more PG information to Zabbix ([pr#22434](#), Wido den Hollander)
- mgr: Miscellaneous small mgr fixes ([pr#22893](#), John Spray)
- mgr: modules CLI commands declaration using @CLICommand decorator ([pr#25543](#), Ricardo Dias)
- mgr,mon: mgr,mon: fix to apply changed mon\_stat\_smooth\_intervals ([pr#23481](#), Yan Jun)
- mgr/orchestrator: added useful attributes to ServiceDescription ([pr#25468](#), Ricardo Dias)
- mgr/orchestrator: Add host mon mgr management to interface ([pr#26314](#), Sebastian Wagner, Noah Watkins)
- mgr/orchestrator: Add JSON output to CLI commands ([pr#25340](#), Sebastian Wagner)
- mgr: orchestrator: add the ability to remove services ([pr#25366](#), Jeff Layton)
- mgr/orchestrator: Allow the orchestrator to scale the NFS server count ([pr#26633](#), Jeff Layton)
- mgr/orchestrator: clarify error message about kubernetes python module ([pr#24525](#), Jeff Layton)
- mgr/orchestrator\_cli: Fix README.md ([pr#26443](#), Sebastian Wagner)
- mgr/orchestrator: Extend DriveGroupSpec ([pr#25912](#), Sebastian Wagner)
- mgr/orchestrator: fix device pretty print with None attributes ([pr#26357](#), Ricardo Dias)
- mgr/orchestrator: fix \_list\_services display ([pr#25610](#), Jeff Layton)
- mgr/orchestrator: Fix up rook osd create dispatcher ([pr#26317](#), Jeff Layton)
- mgr/orchestrator: make use of @CLICommand ([pr#26094](#), Sebastian Wagner)
- mgr/orchestrator: remove unicode whitespaces ([pr#25323](#), Sebastian Wagner)
- mgr/orchestrator/rook: allow the creation of OSDs in directories ([pr#26570](#), Jeff Layton)
- mgr/orchestrator: Unify osd create and osd add ([pr#26171](#), Sebastian Wagner)
- mgr/orch: refresh option for inventory query ([pr#26346](#), Noah Watkins)
- mgr: prometheus: added bluestore db and wal/journal devices to

- ceph\_disk\_occupation metric ([issue#36627](#), [pr#24821](#), Konstantin Shalygin)
- mgr: prometheus: Expose number of degraded/misplaced/unfound objects ([pr#21793](#), Boris Ranto)
- mgr: prometheus: Fix metric resets ([pr#22732](#), Boris Ranto)
- mgr: prometheus: Fix prometheus shutdown/restart ([pr#21748](#), Boris Ranto)
- mgr: pybind/mgr: add osd space utilization to insights report ([pr#25122](#), Noah Watkins)
- mgr: pybind/mgr: PEP 8 code clean and fix typo ([pr#26181](#), Lei Liu)
- mgr,pybind: mgr/prometheus: add interface and objectstore to osd metadata ([pr#25234](#), Jan Fajerski)
- mgr: pybind/mgr/restful: Decode the output of b64decode ([issue#38522](#), [pr#26712](#), Brad Hubbard)
- mgr,pybind: mgr/rook: fix urljoin import ([pr#24626](#), Jeff Layton)
- mgr,pybind: mgr/volumes: Fix Python 3 import error ([pr#25344](#), Sebastian Wagner)
- mgr,pybind: pybind/mgr: drop unnecessary iterkeys usage to make py-3 compatible ([issue#37581](#), [pr#25457](#), Mykola Golub)
- mgr,pybind: pybind/mgr: identify invalid fs ([pr#24392](#), Jos Collin)
- mgr,pybind: src/script: add run\_mypy to run static type checking on Python code ([pr#26715](#), Sebastian Wagner)
- mgr: race between daemon state and service map in ‘service status’ ([issue#36656](#), [pr#24878](#), Mykola Golub)
- mgr,rbd: mgr/prometheus: provide RBD stats via osd dynamic perf counters ([pr#25358](#), Mykola Golub)
- mgr,rbd: pybind/mgr/prometheus: improve ‘rbd\_stats\_pools’ param parsing ([pr#25860](#), Mykola Golub)
- mgr,rbd: pybind/mgr/prometheus: rbd stats namespace support ([pr#25636](#), Mykola Golub)
- mgr: replace “Unknown error” string on always\_on ([pr#23645](#), John Spray)
- mgr: restful: Fix regression when traversing leaf nodes ([pr#26421](#), Boris Ranto)
- mgr/rook: remove dead code and fix bug in url fetching code ([pr#26032](#), Jeff Layton)
- mgr: silence GCC warning ([pr#25199](#), Kefu Chai)

- mgr/ssh: fix type and doc errors ([pr#26630](#), Sebastian Wagner)
- mgr/telemetry: fix total\_objects ([issue#37976](#), [pr#26046](#), Sage Weil)
- mgr,tests: mgr/dashboard: use dedicated tox working dir ([pr#25290](#), Noah Watkins)
- mgr,tests: mgr/insights: use dedicated tox working dir ([pr#25146](#), Noah Watkins)
- mgr,tests: mgr/selftest: fix disabled module selection ([pr#24517](#), John Spray)
- mgr: timely health updates between monitor and manager ([pr#23294](#), Noah Watkins)
- mgr: update daemon\_state when necessary ([issue#37753](#), [pr#25725](#), Xinying Song)
- mgr: update MMgrConfigure message to include optional OSD perf queries ([pr#24180](#), Julien Collet)
- mgr: Use Py\_BuildValue to create the argument tuple ([pr#26240](#), Volker Theile)
- mgr: volumes mgr module fixes ([pr#25331](#), Jeff Layton)
- misc: mark functions with 'override' specifier ([pr#21790](#), Danny Al-Gaaf)
- mon: add 'osd destroy-new' command that only destroys NEW osd slots ([issue#24428](#), [pr#22429](#), Sage Weil)
- mon: A PG with PG\_STATE\_REPAIR doesn't mean damaged data, PG\_STATE\_IN... ([issue#38070](#), [pr#26178](#), David Zafman)
- mon: change monitor compact command to run asynchronously ([issue#24160](#), [issue#24159](#), [pr#22056](#), penglaiyxy)
- mon: common/cmdparse: cmd\_getval\_throws -> cmd\_getval ([pr#23557](#), Sage Weil)
- mon: don't commit osdmap on no-op application ops ([pr#23528](#), John Spray)
- mon: fix mgr module config option handling ([issue#35076](#), [pr#23846](#), Sage Weil)
- mon: fix pg\_sum\_old not copied correctly ([pr#26110](#), Yao Zongyou)
- monitoring/grafana: Fix OSD Capacity Utilization Grafana graph ([pr#24426](#), Maxime)
- mon: make rank ordering explicit (not tied to mon address sort order) ([pr#22193](#), Sage Weil)
- mon: mon/config-key: increase max key entry size ([pr#24250](#), Joao Eduardo Luis)
- mon: mon/MonClient: drop my\_addr ([pr#26449](#), Kefu Chai)
- mon: mon/MonClient: use mon\_client\_ping\_timeout during ping\_monitor ([pr#23563](#), Yao Zongyou)
- mon: mon/MonMap: add more const'ness to its methods ([pr#23709](#), Kefu Chai)

- mon: mon/MonMap: remove duplicate code in get\_rank ([pr#23547](#), Yao Zongyou)
- mon: mon,osd: avoid str copy in parse ([pr#25640](#), Jos Collin)
- mon: mon/OSDMonitor: add boundary check for pool recovery\_priority ([issue#38578](#), [pr#26729](#), xie xingguo)
- mon: mon/PGMap: add more #include ([pr#26420](#), Kefu Chai)
- mon: mon/PGMap: command 'ceph df -f json' output add total\_percent\_used ([pr#23588](#), Yanhu Cao)
- mon: only share monmap after authenticating ([pr#23741](#), Sage Weil)
- mon: shutdown messenger early to avoid accessing deleted logger ([issue#37780](#), [pr#25760](#), ningtao)
- mon: some tiny cleanups related class forward declaration ([pr#26219](#), Yao Zongyou)
- mon,tests: qa/cephtool: test bounds on pool's hit\_set\_\* ([pr#24858](#), Joao Eduardo Luis)
- mon:validate hit\_set values before set ([issue#22659](#), [pr#19983](#), lijing)
- msg: addr -> addrvec (part 1) ([pr#22306](#), Sage Weil)
- msg/async: do not force updating rotating keys inline ([pr#25859](#), yanjun, xie xingguo)
- msg/async/Protocol\*: send keep alive if existing wins ([issue#38493](#), [pr#26668](#), xie xingguo)
- msg/async/rdma: add iWARP RDMA protocol support ([pr#20297](#), Haodong Tang)
- msg/async/rdma: Delete duplicate header file ([pr#25392](#), Jianpeng Ma)
- msg/async/rdma: parse IBSYNMsg.lid as hex when receiving message ([pr#26525](#), Peng Liu)
- msg/async: reduce additional ceph\_msg\_header copy ([pr#25938](#), Jianpeng Ma)
- msg/async: the ceph\_abort is needless in handle\_connect\_msg ([pr#21751](#), shangfufei)
- msg: ceph\_abort() when there are enough accepter errors in msg server ([issue#23649](#), [pr#23306](#), penglaiyxy@gmail.com)
- msg: clear message middle when clearing encoded message buffer ([pr#24289](#), "Yan, Zheng")
- msg: entity\_addr\_t::parse doesn't do memset(this, 0, ...) for clean-up ([issue#26937](#), [pr#23573](#), Radoslaw Zarzynski)

- nautilus: mgr/dashboard: Validate ceph-iscsi config version ([pr#26951](#), Ricardo Marques)
- objecter: avoid race when reset down osd's session ([pr#25437](#), Zengran Zhang)
- orchestrator\_cli: fix HandleCommandResult invocations in \_status() ([pr#25329](#), Jeff Layton)
- osd: add creating to pg\_string\_state ([issue#36174](#), [pr#24262](#), Dan van der Ster)
- osd: add -dump-journal option in ceph-osd help info ([pr#24969](#), yuliyang)
- osd: Additional fields for osd "bench" command ([pr#21962](#), Коренберг Марк)
- osd: add log when pg reg next scrub ([pr#23690](#), lvshuhua)
- osd: add required cls libraries as dependencies of osd ([pr#24373](#), Mohamad Gebai)
- osd: Allow repair of an object with a bad data\_digest in object\_info on all replicas ([pr#23217](#), David Zafman)
- osd: always set query\_epoch explicitly for MOSDPGLog ([pr#22487](#), Kefu Chai)
- osd: avoid using null agent\_state ([pr#25393](#), Zengran Zhang)
- osd: Change assert() to ceph\_assert() missed in the transition ([pr#23918](#), David Zafman)
- osd: Change osd\_skip\_data\_digest default to false and make it LEVEL\_DEV ([issue#24950](#), [pr#23083](#), Sage Weil, David Zafman)
- osdc: invoke notify finish context on linger commit failure ([issue#23966](#), [pr#21831](#), Kefu Chai, Jason Dillaman)
- osd: clean up and avoid extra ref-counting in PrimaryLogPG::log\_op\_stats ([pr#23016](#), Radoslaw Zarzynski)
- osd: clean up smart probe ([issue#23899](#), [pr#21950](#), Sage Weil, Gu Zhongyan)
- osd: collect client perf stats when query is enabled ([pr#24265](#), Julien Collet, Mykola Golub)
- osd: combine recovery/scrub/snap sleep timer into one ([pr#21711](#), Jianpeng Ma)
- osd: Deny reservation if expected backfill size would put us over bac... ([issue#24801](#), [issue#19753](#), [pr#22797](#), David Zafman)
- osd: do not include Messenger.h if not necessary ([pr#22483](#), Kefu Chai)
- osd: do not overestimate the size of the object for reads with trimtrunc ([issue#21931](#), [issue#22330](#), [pr#24564](#), Neha Ojha)

- osd: do not treat an IO hint as an IOP for PG stats ([issue#24909](#), [pr#23029](#), Jason Dillaman)
- osd: don't check overwrite flag when handling copy-get ([issue#21756](#), [pr#18241](#), huangjun)
- osd: Don't evict even when preemption has restarted with smaller chunk ([pr#21892](#), David Zafman)
- osd: do\_sparse\_read(): Verify checksum earlier so we will try to repair ([issue#24875](#), [pr#23377](#), David Zafman)
- osd: drop the unused request\_redirect\_t::osd\_instructions ([pr#24458](#), Radoslaw Zarzynski)
- osd: ec saves a write access to the memory under most circumstances ([pr#26053](#), Zengran Zhang, Kefu Chai)
- osd: fix build\_incremental\_map\_msg ([issue#38282](#), [pr#26413](#), Sage Weil)
- osd: fix memory leak in EC fast and error read ([pr#22500](#), xiaofei cui)
- osd: Fix recovery and backfill priority handling ([issue#38041](#), [pr#26213](#), David Zafman)
- osd: fix shard\_info\_wrapper encode ([issue#37653](#), [pr#25548](#), David Zafman)
- osd: Handle omap and data digests independently ([issue#24366](#), [pr#22346](#), David Zafman)
- osd: increase default hard pg limit ([pr#22187](#), Josh Durgin)
- osd: keep using cache even if op will invalid cache ([pr#25490](#), Zengran Zhang)
- osd: limit pg log length under all circumstances ([pr#23098](#), Neha Ojha)
- osd: make OSD::HEARTBEAT\_MAX\_CONN inline ([pr#23424](#), Kefu Chai)
- osd: make random shuffle comply with C++17 ([pr#23533](#), Willem Jan Withagen)
- osd/OSDMap: add osd status to utilization dumper ([issue#35544](#), [pr#23921](#), Paul Emmerich)
- osd: per-pool osd stats collection ([pr#19454](#), Igor Fedotov, Igor Fedotov)
- osd: Prevent negative local num\_bytes sent to peer for backfill reser... ([issue#38344](#), [pr#26465](#), David Zafman)
- osd: read object attrs failed at EC recovery ([pr#22196](#), xiaofei cui)
- osd: refuse to start if we're > N+2 from recorded require\_osd\_release ([issue#38076](#), [pr#26177](#), Sage Weil)

- osd: reliably send pg\_created messages to the mon ([issue#37775](#), [pr#25731](#), Sage Weil)
- osd: Remove old bft= which has been superceded by backfill ([issue#36170](#), [pr#24256](#), David Zafman)
- osd: remove stray derr ([pr#24042](#), Sage Weil)
- osd: remove unused class read\_log\_and\_missing\_error ([pr#26057](#), Yao Zongyou)
- osd: remove unused fields ([pr#26021](#), Jianpeng Ma)
- osd: remove unused function ([pr#26223](#), Jianpeng Ma)
- osd: Remove useless conditon ([pr#21766](#), Jianpeng Ma)
- osd: some recovery improvements and cleanups ([pr#23663](#), xie xingguo)
- osd: two heartbeat fixes ([pr#25126](#), xie xingguo)
- osd: unlock osd\_lock when tweaking osd settings ([issue#37751](#), [pr#25726](#), Kefu Chai)
- osd: umount store after service.shutdown() ([issue#37975](#), [pr#26043](#), Kefu Chai)
- osd: Weighted Random Sampling for dynamic perf stats ([pr#25582](#), Mykola Golub)
- osd: When possible check CRC in build\_push\_op() so repair can eventually stop ([issue#25084](#), [pr#23518](#), David Zafman)
- osd: write "bench" output to stdout ([issue#24022](#), [pr#21905](#), John Spray)
- os: Minor fixes in comments describing a transaction ([pr#22329](#), Bryan Stillwell)
- performance: Add performance counters breadcrumb ([pr#22060](#), Ricardo Marques)
- performance: mgr/dashboard: Enable gzip compression ([issue#36453](#), [pr#24727](#), Zack Cerza)
- performance: mgr/dashboard: Replace dashboard service ([issue#36675](#), [pr#24900](#), Zack Cerza)
- performance: msg/async: improve read-prefetch logic ([pr#25758](#), xie xingguo)
- performance: qa/tasks/cbt.py: changes to run on bionic ([pr#22405](#), Neha Ojha)
- performance,rbd: common/Throttle: TokenBucketThrottle: use reference to m\_blockers.front() ([issue#36475](#), [pr#24604](#), Dongsheng Yang)
- performance,rbd: pybind/rbd: optimize rbd\_list2 ([pr#25445](#), Mykola Golub)
- Prevent duplicated rows during async tasks ([pr#22148](#), Ricardo Marques)

- prometheus: Fix order of occupation values ([pr#22149](#), Boris Ranto)
- pybind: do not check MFLAGS ([pr#23601](#), Kefu Chai)
- pybind: pybind/ceph\_daemon: expand the order of magnitude of daemonperf statistics to ZB ([issue#23962](#), [pr#21765](#), Guan yunfei)
- pybind: pybind/rbd: make the code more concise ([pr#23664](#), Zheng Yin)
- pybind,rbd: pybind/rbd: add allow\_shrink=True as a parameter to def resize ([pr#23605](#), Zheng Yin)
- pybind,rbd: pybind/rbd: fix a typo in metadata\_get comments ([pr#26138](#), songweibin)
- pybind,rgw: pybind/rgw: pass the flags to callback function ([pr#25766](#), Kefu Chai)
- pybind: simplify timeout handling in run\_in\_thread() ([pr#24733](#), Kefu Chai)
- qa/btrfs/test\_rmdir\_async\_snap: remove binary file ([pr#24108](#), Cleber Rosa)
- qa,pybind,tools: Correct usage of collections.abc ([pr#25318](#), James Page)
- qa/test: Added rados, rbd and fs to run two time a week only ([pr#21839](#), Yuri Weinstein)
- qa/tests: added 1st draft of mimic-x suite ([pr#23292](#), Yuri Weinstein)
- qa/tests - added all supported distro ([pr#22647](#), Yuri Weinstein)
- qa/tests - added all supported distro to the mix, ... ([pr#22674](#), Yuri Weinstein)
- qa/tests: added client-upgrade-luminous suit ([pr#21947](#), Yuri Weinstein)
- qa/tests: added -filter-out="ubuntu\_14.04" ([pr#21949](#), Yuri Weinstein)
- qa/tests - added luminous-p2p suite to the schedule ([pr#22666](#), Yuri Weinstein)
- qa/tests: added mimic-x to the schedule ([pr#23302](#), Yuri Weinstein)
- qa/tests - added powercycle suite to run on weekly basis on master and mimic ([pr#22606](#), Yuri Weinstein)
- qa/tests: added supported distro for powercycle suite ([pr#22185](#), Yuri Weinstein)
- qa/tests: changed ceph qa email address to bypass dreamhost's spam filter ([pr#23456](#), Yuri Weinstein)
- qa/tests: changed disto symlink to point to new way using supported OS'es ([pr#22536](#), Yuri Weinstein)
- qa/tests: fixed typo ([pr#21858](#), Yuri Weinstein)

- qa/tests: removed all jewel runs and reduced runs on ovh ([pr#22531](#), Yuri Weinstein)
- rbd: add ‘config global’ command to get/store overrides in mon config db ([pr#24428](#), Mykola Golub)
- rbd: add data pool support to trash purge ([issue#22872](#), [pr#21247](#), Mahati Chamarthy)
- rbd: add group snap rollback method ([issue#23550](#), [pr#23896](#), songweibin)
- rbd: add protected in snap list ([pr#23853](#), Zheng Yin)
- rbd: add snapshot count in rbd info ([pr#21292](#), Zheng Yin)
- rbd: add the judgment of resizing the image ([pr#21770](#), zhengyin)
- rbd: basic support for images within namespaces ([issue#24558](#), [pr#22673](#), Jason Dillaman)
- rbd: close image when bench is interrupted ([pr#26693](#), Mykola Golub)
- rbd: cls/lock: always store v1 addr in locker\_info\_t ([pr#25948](#), Sage Weil)
- rbd: cls/rbd: fix build ([pr#22078](#), Kefu Chai)
- rbd: cls/rbd: fixed uninitialized variable compiler warning ([pr#26896](#), Jason Dillaman)
- rbd: cls/rbd: fix method comment ([pr#23277](#), Zheng Yin)
- rbd: cls/rbd: silence the log of get metadata error ([pr#25436](#), songweibin)
- rbd: correct parameter of namespace and verify it before set\_namespace ([pr#23770](#), songweibin)
- rbd: dashboard: support configuring block mirroring pools and peers ([pr#25210](#), Jason Dillaman)
- rbd: disable cache for actions that open multiple images ([issue#24092](#), [pr#21946](#), Jason Dillaman)
- rbd: disk-usage can now optionally compute exact on-disk usage ([issue#24064](#), [pr#21912](#), Jason Dillaman)
- rbd: Document new RBD feature flags and version support ([pr#25192](#), Valentin Lorentz)
- rbd: don’t load config overrides from monitor initially ([pr#21910](#), Jason Dillaman)
- rbd: error if new size is equal to original size ([pr#22637](#), zhengyin)

- rbd: expose pool stats summary tool ([pr#24830](#), Jason Dillaman)
- rbd: filter out group/trash snapshots from snap\_list ([pr#23638](#), songweibin)
- rbd: fix a typo in error output ([pr#25931](#), Dongsheng Yang)
- rbd: fix delay time calculation for trash move ([pr#25896](#), Mykola Golub)
- rbd: fix error import when the input is a pipe ([issue#34536](#), [pr#23835](#), songweibin)
- rbd: fix segmentation fault when rbd\_group\_image\_list() getting -ENOENT ([issue#38468](#), [pr#26622](#), songweibin)
- rbd: fix some typos ([pr#25083](#), Shiyang Ruan)
- rbd: implement new 'rbd perf image iostat/iotop' commands ([issue#37913](#), [pr#26133](#), Jason Dillaman)
- rbd: improved trash snapshot namespace handling ([issue#23398](#), [pr#23191](#), Jason Dillaman)
- rbd: interlock object-map/fast-diff features together ([pr#21969](#), Mao Zhongyi)
- rbd: introduce abort\_on\_full option for rbd map ([pr#25662](#), Dongsheng Yang)
- rbd: journal: allow remove set when journal pool is full ([pr#25166](#), kungf)
- rbd: journal: fix potential race when closing object recorder ([pr#26425](#), Mykola Golub)
- rbd: journal: set max journal order to 26 ([issue#37541](#), [pr#25743](#), Mykola Golub)
- rbd: krbd: support for images within namespaces ([pr#23841](#), Ilya Dryomov)
- rbd: librbd/api: misc fix migration ([pr#25765](#), songweibin)
- rbd: librbd: ensure exclusive lock acquired when removing sync point snapshots ([issue#24898](#), [pr#23095](#), Mykola Golub)
- rbd: librbd: misc fix potential invalid pointer ([pr#25462](#), songweibin)
- rbd: make sure the return-value 'r' will be returned ([pr#24891](#), Shiyang Ruan)
- rbd: mgr/dashboard: incorporate RBD overall performance grafana dashboard ([issue#37867](#), [pr#25927](#), Jason Dillaman)
- rbd-mirror: always attempt to restart canceled status update task ([issue#36500](#), [pr#24646](#), Jason Dillaman)
- rbd-mirror: bootstrap needs to handle local image id collision ([issue#24139](#), [pr#22043](#), Jason Dillaman)

- rbd-mirror: create and export replication perf counters to mgr ([pr#25834](#), Mykola Golub)
- rbd-mirror: ensure daemon can cleanly exit if pool is deleted ([pr#22348](#), Jason Dillaman)
- rbd-mirror: ensure remote demotion is replayed locally ([issue#24009](#), [pr#21823](#), Jason Dillaman)
- rbd-mirror: fixed potential crashes during shut down ([issue#24008](#), [pr#21817](#), Jason Dillaman)
- rbd-mirror: guard access to image replayer perf counters ([pr#26097](#), Mykola Golub)
- rbd-mirror: instantiate the status formatter before changing state ([issue#36084](#), [pr#24181](#), Jason Dillaman)
- rbd-mirror: optionally extract peer secrets from config-key ([issue#24688](#), [pr#24036](#), Jason Dillaman)
- rbd-mirror: optionally support active/active replication ([pr#21915](#), Mykola Golub, Jason Dillaman)
- rbd-mirror: potential deadlock when running asok 'flush' command ([issue#24141](#), [pr#22027](#), Mykola Golub)
- rbd-mirror: prevent creation of clones when parents are syncing ([issue#24140](#), [pr#24063](#), Jason Dillaman)
- rbd-mirror: schedule rebalancer to level-load instances ([issue#24161](#), [pr#22304](#), Venky Shankar)
- rbd-mirror: update mirror status when stopping ([issue#36659](#), [pr#24864](#), Jason Dillaman)
- rbd-mirror: use active/active policy by default ([issue#38453](#), [pr#26603](#), Jason Dillaman)
- rbd: move image to trash as first step when removing ([issue#24226](#), [issue#38404](#), [pr#25438](#), Mahati Chamarthy, Jason Dillaman)
- rbd-nbd: do not ceph\_abort() after print the usages ([issue#36660](#), [pr#24815](#), Shiyang Ruan)
- rbd-nbd: support namespaces ([issue#24609](#), [pr#25260](#), Mykola Golub)
- rbd: not allowed to restore an image when it is being deleted ([issue#25346](#), [pr#24078](#), songweibin)
- rbd: online re-sparsify of images ([pr#26226](#), Mykola Golub)
- rbd: pybind/rbd: add namespace helper API methods ([issue#36622](#), [pr#25206](#), Jason

Dillaman)

- rbd: qa/workunits: fixed mon address parsing for rbd-mirror ([issue#38385](#), [pr#26521](#), Jason Dillaman)
- rbd: rbd: fix error parse arg when getting key ([pr#25152](#), songweibin)
- rbd: rbd-fuse: look for ceph.conf in standard locations ([issue#12219](#), [pr#20598](#), Jason Dillaman)
- rbd: rbd-fuse: namespace support ([pr#25265](#), Mykola Golub)
- rbd: rbd-ggate: support namespaces ([issue#24608](#), [pr#25266](#), Mykola Golub)
- rbd: rbd-ggate: tag "level" with need\_dynamic ([pr#22557](#), Kefu Chai)
- rbd: rbd\_mirror: assert no requests on destroying InstanceWatcher ([pr#25666](#), Mykola Golub)
- rbd: rbd\_mirror: don't report error if image replay canceled ([pr#25789](#), Mykola Golub)
- rbd: rbd-mirror: use pool level config overrides ([pr#24348](#), Mykola Golub)
- rbd: rbd: show info about mirror daemon instance in image mirror status output ([pr#24717](#), Mykola Golub)
- rbd: return error code when the source and distination namespace are different ([pr#24893](#), Shiyang Ruan)
- rbd: simplified code to remove do\_clear\_limit function ([pr#23954](#), Zheng Yin)
- rbd: support namespaces for image migration ([issue#26951](#), [pr#24836](#), Jason Dillaman)
- rbd: systemd/rbdmap.service: order us before remote-fs-pre.target ([issue#24713](#), [pr#22769](#), Ilya Dryomov)
- rbd: test/librbd: drop unused variable 'num\_aios' ([pr#23085](#), songweibin)
- rbd,tests: krbd: alloc\_size map option and tests ([pr#26244](#), Ilya Dryomov)
- rbd,tests: librbd,test: remove unused context\_cb() function, silence GCC warnings ([pr#24673](#), Kefu Chai)
- rbd,tests: pybind/rbd: add assert\_raise in test set\_snap ([pr#22570](#), Zheng Yin)
- rbd,tests: qa: krbd\_exclusive\_option.sh: bump lock\_timeout to 60 seconds ([issue#25080](#), [pr#22648](#), Ilya Dryomov)
- rbd,tests: qa: krbd\_msgr\_segments.t: filter lvcreate output ([pr#22665](#), Ilya Dryomov)

- rbd,tests: qa: krbd namespaces test ([pr#26339](#), Ilya Dryomov)
- rbd,tests: qa: objectstore snippets for krbd ([pr#26279](#), Ilya Dryomov)
- rbd,tests: qa: rbd\_workunit\_kernel\_untar\_build: install build dependencies ([issue#35074](#), [pr#23840](#), Ilya Dryomov)
- rbd,tests: qa: rbd/workunits : Replace “rbd bench-write” with “rbd bench -io-type write” ([pr#26168](#), Shyukri Shyukriev)
- rbd,tests: qa/suites/krbd: more fsx tests ([pr#24354](#), Ilya Dryomov)
- rbd,tests: qa/suites/rbd: randomly select a supported distro ([pr#22008](#), Jason Dillaman)
- rbd,tests: qa/tasks/cram: tasks now must live in the repository ([pr#23976](#), Ilya Dryomov)
- rbd,tests: qa/tasks/cram: use suite\_repo repository for all cram jobs ([pr#23905](#), Ilya Dryomov)
- rbd,tests: qa/tasks/qemu: use unique clone directory to avoid race with workunit ([issue#36542](#), [pr#24696](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: fix cli generic namespace test ([pr#24457](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: force v2 image format for namespace test ([pr#24512](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: replace usage of ‘rados mkpool’ ([pr#23938](#), Jason Dillaman)
- rbd,tests: qa/workunits: replace ‘realpath’ with ‘readlink -f’ in fsstress.sh ([issue#36409](#), [pr#24550](#), Jason Dillaman)
- rbd,tests: test/cli-integration/rbd: added new parent image attributes ([pr#25415](#), Jason Dillaman)
- rbd,tests: test/librados\_test\_stub: deterministically load cls shared libraries ([pr#21524](#), Jason Dillaman)
- rbd,tests: test/librados\_test\_stub: handle object doesn’t exist gracefully ([pr#25667](#), Mykola Golub)
- rbd,tests: test/librbd: fix compiler -Wsign-compare warnings ([pr#23657](#), Mykola Golub)
- rbd,tests: test/librbd: fix gmock warning in snapshot rollback test ([pr#23736](#), Jason Dillaman)
- rbd,tests: test/librbd: fix gmock warning in

- TestMockIoImageRequestWQ.AcquireLockError ([pr#22778](#), Mykola Golub)
- rbd,tests: test/librbd: fix gmock warnings for get\_modify\_timestamp call ([pr#23707](#), Mykola Golub)
- rbd,tests: test/librbd: fix 'Uninteresting mock function call' warning ([pr#26322](#), Mykola Golub)
- rbd,tests: test/librbd: fix valgrind warnings ([pr#23827](#), Mykola Golub)
- rbd,tests: test/librbd: fix -Wsign-compare warnings ([pr#23608](#), Kefu Chai)
- rbd,tests: test/librbd: metadata key for config should be prefixed with `conf_` ([pr#25209](#), runsisi)
- rbd,tests: test/librbd: migration supporting namespace tests ([pr#24919](#), Mykola Golub)
- rbd,tests: test/librbd: migration tests did not delete additional pool ([pr#24009](#), Mykola Golub)
- rbd,tests: test: move OpenStack devstack test to rocky release ([issue#36410](#), [pr#24563](#), Jason Dillaman)
- rbd,tests: test/pybind: fix test\_rbd.TestClone.test\_trash\_snapshot ([issue#25114](#), [pr#23256](#), Mykola Golub)
- rbd,tests: test/pybind/test\_rbd: filter out unknown list\_children2 keys ([issue#37729](#), [pr#25832](#), Mykola Golub)
- rbd,tests: test/rbd-mirror: disable use of gtest-parallel ([pr#22694](#), Jason Dillaman)
- rbd,tests: test/rbd\_mirror: fix gmock warnings ([pr#25863](#), Mykola Golub)
- rbd,tests: test/rbd\_mirror: race in TestMockImageMap.AddInstancePingPongImageTest ([issue#36683](#), [pr#24897](#), Mykola Golub)
- rbd,tests: test/rbd\_mirror: race in WaitingOnLeaderReleaseLeader ([issue#36236](#), [pr#24300](#), Mykola Golub)
- rbd,tests: test/rbd\_mirror: wait for release leader lock fully complete ([pr#25935](#), Mykola Golub)
- rbd,tests: test/rbd: rbd\_ggate test improvements ([pr#23630](#), Willem Jan Withagen)
- rbd,tests: test: silence -Wsign-compare warnings ([pr#23655](#), Kefu Chai)
- rbd: tools/rbd/action: align column headers left ([pr#22566](#), Sage Weil)
- rbd: tools/rbd: assert(g\_ceph\_context) not g\_conf ([pr#23167](#), Kefu Chai)

- rbd: tools/rbd: minor fixes for rbd du display ([pr#23311](#), songweibin)
- rbd,tools: rbd-mirror,common: fix typos in logging messages and comments ([pr#25197](#), Shiyang Ruan)
- rbd,tools: tools/rbd: assert(g\_ceph\_context) not g\_conf ([pr#23008](#), Kefu Chai)
- rbd: wait for all io complete when bench is interrupted ([pr#26918](#), Mykola Golub)
- rbd: workaround for llvm linker problem, avoid std::pair dtor ([pr#25301](#), Willem Jan Withagen)
- Revert "cephfs-journal-tool: enable purge\_queue journal's event comma... ([pr#23465](#), "Yan, Zheng")
- Revert "ceph-fuse: Delete inode's bufferhead was in Tx state would le... ([pr#21975](#), "Yan, Zheng")
- rgw: abort\_bucket\_multiparts() ignores individual NoSuchUpload errors ([issue#35986](#), [pr#24110](#), Casey Bodley)
- rgw: adapt AioThrottle for RGWGetObj ([pr#25208](#), Casey Bodley)
- rgw: Add append object api ([pr#22755](#), zhang Shaowen, Zhang Shaowen)
- rgw: add bucket as option when show/trim usage ([pr#23819](#), lvshuhua)
- rgw: add configurable AWS-compat invalid range get behavior ([issue#24317](#), [pr#22231](#), Matt Benjamin)
- rgw: add curl\_low\_speed\_limit and curl\_low\_speed\_time config to avoid ([pr#23058](#), Mark Kogan, Zhang Shaowen)
- rgw: add Http header 'Server' in response headers ([pr#23282](#), Zhang Shaowen)
- rgw: Adding documentation for Roles ([pr#24714](#), Pritha Srivastava)
- rgw: add latency info in the log of req done ([pr#23906](#), lvshuhua)
- rgw: add list user admin OP API ([pr#25073](#), Oshyn Song)
- rgw: add -op-mask in radosgw-admin help info ([pr#24848](#), yuliyang)
- rgw: add optional\_yield to block\_while\_resharding() ([pr#25357](#), Casey Bodley)
- rgw: add option for relaxed region enforcement ([issue#24507](#), [pr#22533](#), Matt Benjamin)
- rgw: Add rgw xml unit tests ([pr#26682](#), Yuval Lifshitz)
- rgw: add s3 notification sub resources ([pr#23405](#), yuliyang)
- rgw: admin rest api support op-mask ([pr#24869](#), yuliyang)

- rgw: admin/user ops dump user 'system' flag ([pr#17414](#), fang.yuxiang)
- rgw: All Your Fault ([issue#24962](#), [pr#23099](#), Adam C. Emerson)
- rgw: apply quota config to users created via external auth ([issue#24595](#), [pr#24177](#), Casey Bodley)
- rgw: archive zone ([pr#25137](#), Yehuda Sadeh, Javier M. Mellid)
- rgw: async sync\_object and remove\_object does not access coroutine me... ([issue#35905](#), [pr#24007](#), Tianshan Qu)
- rgw: async watch registration ([pr#21838](#), Yehuda Sadeh)
- rgw: avoid race condition in RGWHTTPClient::wait() ([pr#21767](#), cfanz)
- rgw: beast frontend logs socket errors at level 4 ([pr#24677](#), Casey Bodley)
- rgw: beast frontend parses ipv6 addrs ([issue#36662](#), [pr#24887](#), Casey Bodley)
- rgw: beast frontend reworks pause/stop and yields during body io ([pr#21271](#), Casey Bodley)
- rgw: bucket full sync handles delete markers ([issue#38007](#), [pr#26081](#), Casey Bodley)
- rgw: bucket limit check misbehaves for > max-entries buckets (usually... ([pr#26800](#), Matt Benjamin)
- rgw: bucket sync status improvements, part 1 ([pr#21788](#), Casey Bodley)
- rgw: bug in versioning concurrent, list and get have consistency issue ([pr#26197](#), Wang Hao)
- rgw: catch exceptions from librados::NObjectIterator ([issue#37091](#), [pr#25081](#), Casey Bodley)
- rgw: change default rgw\_thread\_pool\_size to 512 ([issue#24544](#), [pr#22581](#), Douglas Fuller)
- rgw: change the "rgw admin status" 'num\_shards' output to signed int ([issue#37645](#), [pr#25538](#), Mark Kogan)
- rgw: check for non-existent bucket in RGWGetACLs ([pr#26212](#), Matt Benjamin)
- rgw: civetweb: update for url validation fixes ([issue#24158](#), [pr#22054](#), Abhishek Lekshmanan)
- rgw: civetweb: use poll instead of select while waiting on sockets ([issue#24364](#), [pr#24027](#), Abhishek Lekshmanan)
- rgw: clean-up - insure C++ source code files contain editor directives ([pr#25495](#),

J. Eric Ivancich)

- rgw: cleanups for sync tracing ([pr#23828](#), Casey Bodley)
- rgw: clean-up - use enum class for stats category ([pr#25450](#), J. Eric Ivancich)
- rgw: cls/rgw: don't assert in decode\_list\_index\_key() ([issue#24117](#), [pr#22440](#), Yehuda Sadeh)
- rgw: cls/rgw: raise debug level of bi\_log\_iterate\_entries output ([pr#25570](#), Casey Bodley)
- rgw: cls/user: cls\_user\_remove\_bucket writes modified header ([issue#36496](#), [pr#24645](#), Casey Bodley)
- rgw: Code for STS Authentication ([pr#23504](#), Pritha Srivastava)
- rgw: common/options: correct the description of rgw\_enable\_lc\_threads option ([pr#23511](#), excellentkf)
- rgw: continue enoent index in dir\_suggest ([issue#24640](#), [pr#22937](#), Tianshan Qu)
- rgw: copy actual stats from the source shards during reshard ([issue#36290](#), [pr#24444](#), Abhishek Lekshmanan)
- rgw: Copying object data should generate new tail tag for the new object ([issue#24562](#), [pr#22613](#), Zhang Shaowen)
- rgw: Correcting logic for signature calculation for non s3 ops ([pr#26098](#), Pritha Srivastava)
- rgw: cors rules num limit ([pr#23434](#), yuliyang)
- rgw: crypto: add openssl support for RGW encryption ([pr#15168](#), Qiaowei Ren)
- rgw: data sync accepts ERR\_PRECONDITION\_FAILED on remove\_object() ([issue#37448](#), [pr#25310](#), Casey Bodley)
- rgw: data sync drains lease stack on lease failure ([issue#38479](#), [pr#26639](#), Casey Bodley)
- rgw: data sync respects error\_retry\_time for backoff on error\_repo ([issue#26938](#), [pr#23571](#), Casey Bodley)
- rgw: delete multi object num limit ([pr#23544](#), yuliyang)
- rgw: delete some unused code about std::regex ([pr#23221](#), Xueyu Bai)
- rgw: [DNM] rgw: Controlling STS authentication via a Policy ([pr#24818](#), Pritha Srivastava)
- rgw: do not ignore EEXIST in RGWPutObj::execute ([issue#22790](#), [pr#23033](#), Matt

Benjamin)

- rgw: Do not modify email if argument is not set ([pr#22024](#), Volker Theile)
- rgw: dont access rgw\_http\_req\_data::client of canceled request ([issue#35851](#), [pr#23988](#), Casey Bodley)
- rgw: Don't treat colons specially when matching resource field of ARNs in S3 Policy ([issue#23817](#), [pr#25145](#), Adam C. Emerson)
- rgw: drop unused tmp in main() ([pr#23899](#), luomuyao)
- rgw: escape markers in RGWOp\_Metadata\_List::execute ([issue#23099](#), [pr#22721](#), Matt Benjamin)
- rgw: ES sync: be more restrictive on object system attrs ([issue#36233](#), [pr#24492](#), Abhishek Lekshmanan)
- rgw: etag in rgw copy result response body rather in header ([pr#23751](#), yuliyang)
- rgw: feature - log successful bucket resharding events ([pr#25510](#), J. Eric Ivancich)
- rgw: fetch\_remote\_obj filters out olh attrs ([issue#37792](#), [pr#25794](#), Casey Bodley)
- rgw: fix bad user stats on versioned bucket after reshard ([pr#25414](#), J. Eric Ivancich)
- rgw: fix build ([pr#22194](#), Yehuda Sadeh)
- rgw: fix build ([pr#23248](#), Matt Benjamin)
- rgw: fix chunked-encoding for chunks >1MiB ([issue#35990](#), [pr#24114](#), Robin H. Johnson)
- rgw: fix compilation after pubsub conflict ([pr#25568](#), Casey Bodley)
- rgw: fix copy response header etag format not correct ([issue#24563](#), [pr#22614](#), Tianshan Qu)
- rgw: fix CreateBucket with BucketLocation parameter failed under default zonegroup ([pr#22312](#), Enming Zhang)
- rgw: fix deadlock on RGWIndexCompletionManager::stop ([issue#26949](#), [pr#23590](#), Yao Zongyou)
- rgw: fix dependencies/target\_link\_libraries ([pr#23056](#), Michal Jarzabek)
- rgw: fixes for sync of versioned objects ([issue#24367](#), [pr#22347](#), Casey Bodley)
- rgw: Fixes to permission evaluation related to user policies ([pr#25180](#), Pritha Srivastava)

- rgw: fix Etag error in multipart copy response ([pr#23749](#), yuliyang)
- rgw: Fix for buffer overflow in STS op\_post() ([issue#36579](#), [pr#24510](#), Pritha Srivastava, Marcus Watts)
- rgw: Fix for SignatureMismatchError in s3 commands ([pr#26204](#), Pritha Srivastava)
- rgw: fix FTBFS introduced by abca9805 ([pr#23046](#), Kefu Chai)
- rgw: fix index complete miss zones\_trace set ([issue#24590](#), [pr#22632](#), Tianshan Qu)
- rgw: fix index update in dir\_suggest\_changes ([issue#24280](#), [pr#22217](#), Tianshan Qu)
- rgw: fix ldap secret parsing ([pr#25796](#), Matt Benjamin)
- rgw: fix leak of curl handle on shutdown ([issue#35715](#), [pr#23986](#), Casey Bodley)
- rgw: Fix log level of gc\_iterate\_entries ([issue#23801](#), [pr#22868](#), iliul)
- rgw: fix max-size in radosgw-admin and REST Admin API ([pr#24062](#), Nick Erdmann)
- rgw: fix meta and data notify thread miss stop cr manager ([issue#24589](#), [pr#22631](#), Tianshan Qu)
- rgw: fix obj can still be deleted even if deleteobject policy is set ([issue#37403](#), [pr#25278](#), Enming.Zhang)
- rgw: fix radosgw-admin build error ([pr#21599](#), cfanz)
- rgw: fix rgw\_data\_sync\_info::json\_decode() ([issue#38373](#), [pr#26494](#), Casey Bodley)
- rgw: fix RGWSyncTraceNode crash in reload ([issue#24432](#), [pr#22432](#), Tianshan Qu)
- rgw: fix stats for versioned buckets after reshards ([pr#25333](#), J. Eric Ivancich)
- rgw: fix uninitialized access ([pr#25002](#), Yehuda Sadeh)
- rgw: fix unordered bucket listing when object names are adorned ([issue#38486](#), [pr#26658](#), J. Eric Ivancich)
- rgw: fix vector index out of range in RGWReadDataSyncRecoveringShardsCR ([issue#36537](#), [pr#24680](#), Casey Bodley)
- rgw: fix version bucket stats ([issue#21429](#), [pr#17789](#), Shasha Lu)
- rgw: fix versioned obj copy generating tags ([issue#37588](#), [pr#25473](#), Abhishek Lekshmanan)
- rgw: fix wrong debug related to user ACLs in rgw\_build\_bucket\_policies() ([issue#19514](#), [pr#14369](#), Radoslaw Zarzynski)
- rgw: get or set realm zonegroup zone need check user's caps ([pr#25178](#), yuliyang, Casey Bodley)

- rgw: Get the user metadata of the user used to sign the request ([pr#22390](#), Volker Theile)
- rgw: handle cases around zone deletion ([issue#37328](#), [pr#25160](#), Abhishek Lekshmanan)
- rgw: handle S3 version 2 pre-signed urls with meta-data ([pr#24683](#), Matt Benjamin)
- rgw: have a configurable authentication order ([issue#23089](#), [pr#21494](#), Abhishek Lekshmanan)
- rgw: http client: print curl error messages during curl failures ([pr#23318](#), Abhishek Lekshmanan)
- rgw: Improvements to STS Lite documentation ([pr#24847](#), Pritha Srivastava)
- rgw: Initial commit for AssumeRoleWithWebIdentity ([pr#26002](#), Pritha Srivastava)
- rgw: initial RGWRados refactoring work ([pr#24014](#), Yehuda Sadeh, Casey Bodley)
- rgw: Initial work for OPA-Ceph integration ([pr#22624](#), Ashutosh Narkar)
- rgw: librgw: initialize curl and http client for multisite ([issue#36302](#), [pr#24402](#), Casey Bodley)
- rgw: librgw: support symbolic link ([pr#19684](#), Tao Chen)
- rgw: lifecycle: don't reject compound rules with empty prefix ([issue#37879](#), [pr#25926](#), Matt Benjamin)
- rgw: Limit the number of lifecycle rules on one bucket ([issue#24572](#), [pr#22623](#), Zhang Shaowen)
- rgw: list bucket can not show the object uploaded by RGWPostObj when enable bucket versioning ([pr#24341](#), yuliyang)
- rgw: log http status with op prefix if available ([pr#25102](#), Casey Bodley)
- rgw: log refactoring for data sync ([pr#23843](#), Casey Bodley)
- rgw: log refactoring for meta sync ([pr#23950](#), Casey Bodley, Ali Maredia)
- rgw: make beast the default for rgw\_frontends ([pr#26599](#), Casey Bodley)
- rgw: Minor fixes to AssumeRole for boto compliance ([pr#24845](#), Pritha Srivastava)
- rgw: Minor fixes to radosgw-admin commands for a role ([pr#24730](#), Pritha Srivastava)
- rgw: move all reshards config options out of legacy\_config\_options ([pr#25356](#), J. Eric Ivancich)

- rgw: move keystone secrets from ceph.conf to files ([issue#36621](#), [pr#24816](#), Matt Benjamin)
- rgw: multiple es related fixes and improvements ([issue#22877](#), [issue#38028](#), [issue#38030](#), [issue#36092](#), [pr#26106](#), Yehuda Sadeh, Abhishek Lekshmanan)
- rgw: need to give a type in list constructor ([pr#25161](#), Willem Jan Withagen)
- rgw: new librgw\_admin\_us ([pr#21439](#), Orit Wasserman, Matt Benjamin)
- rgw: policy: fix NotAction, NotPrincipal, NotResource does not take effect ([pr#23625](#), xiangxiang)
- rgw: policy: fix s3:x-amz-grant-read-acp keyword error ([pr#23610](#), xiangxiang)
- rgw: policy: modify some operation permission keyword ([issue#24061](#), [pr#20974](#), xiangxiang)
- rgw: pub-sub ([pr#23298](#), Yehuda Sadeh)
- rgw: qa/suites/rgw/verify/tasks/cls\_rgw: test cls\_rgw ([pr#22919](#), Sage Weil)
- rgw: radogw-admin reshards status command should print text for reshards status ([issue#23257](#), [pr#20779](#), Orit Wasserman)
- rgw: radosgw-admin: add mfa related command and options ([pr#23416](#), Enming.Zhang)
- rgw: radosgw-admin bucket rm ... --purge-objects can hang ([issue#38134](#), [pr#26231](#), J. Eric Ivancich)
- rgw: "radosgw-admin objects expire" always returns ok even if the process fails ([issue#24592](#), [pr#22635](#), Zhang Shaowen)
- rgw: radosgw-admin: 'sync error trim' loops until complete ([issue#24873](#), [pr#23032](#), Casey Bodley)
- rgw: radosgw-admin: translate reshards status codes (trivial) ([issue#36486](#), [pr#24638](#), Matt Benjamin)
- rgw: RADOS::Obj::operate takes optional\_yield ([pr#25068](#), Casey Bodley)
- rgw: rados tiering ([issue#19510](#), [pr#25774](#), yuliyang, Yehuda Sadeh, Zhang Shaowen)
- rgw: raise debug level on redundant data sync error messages ([issue#35830](#), [pr#23981](#), Casey Bodley)
- rgw: raise default rgw\_curl\_low\_speed\_time to 300 seconds ([issue#27989](#), [pr#23759](#), Casey Bodley)
- rgw: refactor logging in gc and lc ([pr#24530](#), Ali Maredia)
- rgw: refactor PutObjProcessor stack ([pr#24453](#), Casey Bodley)

- rgw: reject invalid methods in validate\_cors\_rule\_method ([issue#24223](#), [pr#22145](#), Jeegn Chen)
- rgw: remove all traces of cls replica\_log ([pr#21680](#), Casey Bodley)
- rgw: remove duplicated `RGWRados::list_buckets_` helpers ([pr#25240](#), Casey Bodley)
- rgw: remove expired entries from the cache ([issue#23379](#), [pr#22410](#), Mark Kogan)
- rgw: remove repetitive conditional statement in RGWHandler\_REST\_Obj\_S3 ([pr#24162](#), Zhang Shaowen)
- rgw: remove rgw\_aclparser.cc ([issue#36665](#), [pr#24866](#), Matt Benjamin)
- rgw: remove the useless `is_cors_op` in RGWHandler\_REST\_Obj\_S3 ([pr#22114](#), Zhang Shaowen)
- rgw: remove unused aio helper functions ([pr#25239](#), Casey Bodley)
- rgw: renew resharding locks to prevent expiration ([issue#27219](#), [issue#34307](#), [pr#24406](#), Orit Wasserman, J. Eric Ivancich)
- rgw: repair olh attributes that were broken by sync ([issue#37792](#), [pr#26157](#), Casey Bodley)
- rgw: require `-yes-i-really-mean-it` to run radosgw-admin orphans find ([issue#24146](#), [pr#22036](#), Matt Benjamin)
- rgw: reshards add: fail correctly on a non existant bucket ([issue#36449](#), [pr#24594](#), Abhishek Lekshmanan)
- rgw: reshards clean-up and associated commits ([pr#25142](#), J. Eric Ivancich)
- rgw: reshards improvements ([pr#25003](#), J. Eric Ivancich)
- rgw: reshards stale instance cleanup ([issue#24082](#), [pr#24662](#), Abhishek Lekshmanan)
- rgw: resolve bugs and clean up garbage collection code ([issue#38454](#), [pr#26601](#), J. Eric Ivancich)
- rgw: resolve bug where marker was not advanced during garbage collection ([issue#38408](#), [pr#26545](#), J. Eric Ivancich)
- rgw: return `err_malformed_xml` when `MaxAgeSeconds` is an invalid integer ([issue#26957](#), [pr#23626](#), Chang Liu)
- rgw: Return tenant field in `bucket_stats` function ([pr#24895](#), Volker Theile)
- rgw: return valid `Location` element, `PostObj` ([issue#22927](#), [pr#20330](#), yuliyang)
- rgw: return `x-amz-version-id: null` when delete obj in versioning suspended bucket ([issue#35814](#), [pr#23927](#), yuliyang)

- rgw: Revert "rgw: lifecycle: don't reject compound rules with empty prefix" ([pr#26491](#), Matt Benjamin)
- rgw: rgw-admin: add "-trim-delay-ms" introduction for 'sync error trim' ([pr#23342](#), Enming.Zhang)
- rgw: rgw-admin: fix data sync report for master zone ([pr#23925](#), cfanz)
- rgw: RGWAsyncGetBucketInstanceInfo does not access coroutine memory ([issue#35812](#), [pr#23987](#), Casey Bodley)
- rgw: rgw/beast: drop privileges after binding ports ([issue#36041](#), [pr#24271](#), Paul Emmerich)
- rgw: RGWBucket::link supports tenant ([issue#22666](#), [pr#23119](#), Casey Bodley)
- rgw: rgw: change the way sysobj filters raw attributes, fix bucket sync state xattrs ([issue#37281](#), [pr#25123](#), Yehuda Sadeh)
- rgw: rgw, cls: remove cls\_statelog and rgw opstate tracking ([pr#24059](#), Casey Bodley)
- rgw: rgw\_file: deep stat handling ([issue#24915](#), [pr#23038](#), Matt Benjamin)
- rgw: rgw\_file: not check max\_objects when creating file ([pr#24846](#), Tao Chen)
- rgw: rgw\_file: use correct secret key to check auth ([pr#26130](#), MinSheng Lin)
- rgw: rgw\_file: user info never synced since librgw init ([pr#25406](#), Tao Chen)
- rgw: [rgw]: Fix help of radosgw-admin user info in case no uid ([pr#25078](#), Marc Koderer)
- rgw: rgwgc:process coredump in some special case ([issue#23199](#), [pr#25430](#), zhaokun)
- rgw: rgw multisite: async rados requests don't access coroutine memory ([issue#35543](#), [pr#23920](#), Casey Bodley)
- rgw: rgw multisite: bucket sync transitions back to StateInit on OP\_SYNCSTOP ([issue#26895](#), [pr#23574](#), Casey Bodley)
- rgw: rgw multisite: enforce spawn\_window for data full sync ([issue#26897](#), [pr#23534](#), Casey Bodley)
- rgw: rgw-multisite: fix endless loop in RGWBucketShardIncrementalSyncCR ([issue#24603](#), [pr#22660](#), cfanz)
- rgw: rgw multisite: incremental data sync uses truncated flag to detect end of listing ([issue#26952](#), [pr#23596](#), Casey Bodley)
- rgw: rgw multisite: only update last\_trim marker on ENODATA ([issue#38075](#), [pr#26190](#), Casey Bodley)

- rgw: rgw multisite: uses local DataChangesLog to track active buckets for trim ([issue#36034](#), [pr#24221](#), Casey Bodley)
- rgw: rgw/pubsub: add amqp push endpoint ([pr#25866](#), Yuval Lifshitz)
- rgw: rgw/pubsub: add pubsub tests ([pr#26299](#), Yuval Lifshitz)
- rgw: RGWRadosGetOmapKeysCR takes result by shared\_ptr ([issue#21154](#), [pr#23634](#), Casey Bodley)
- rgw: RGWRadosGetOmapKeysCR uses 'more' flag from omap\_get\_keys2() ([pr#23401](#), Casey Bodley, Sage Weil)
- rgw: remove duplicate include header files in rgw\_rados.cc ([pr#18578](#), Sibei Gao)
- rgw: rgw\_sync: drop ENOENT error logs from mdlog ([pr#26971](#), Abhishek Lekshmanan)
- rgw: Robustly notify ([issue#24963](#), [pr#23100](#), Adam C. Emerson)
- rgw: s3: awsv4 drop special handling for x-amz-credential ([issue#26965](#), [pr#23652](#), Abhishek Lekshmanan)
- rgw: sanitize customer encryption keys from log output in v4 auth ([issue#37847](#), [pr#25881](#), Casey Bodley)
- rgw: scheduler ([pr#26008](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: set cr state if aio\_read err return in RGWCloneMetaLogCoroutine ([issue#24566](#), [pr#22617](#), Tianshan Qu)
- rgw: set default objecter\_inflight\_ops = 24576 ([issue#25109](#), [pr#23242](#), Matt Benjamin)
- rgw: should recode canonical\_uri when caculate s3 v4 auth ([issue#23587](#), [pr#21286](#), yuliyang)
- rgw: some fix for es sync ([issue#23842](#), [issue#23841](#), [pr#21622](#), Tianshan Qu, Shang Ding)
- rgw: support admin rest api get user info through user's access-key ([pr#22790](#), yuliyang)
- rgw: support server-side encryption when SSL is terminated in a proxy ([issue#27221](#), [pr#24700](#), Casey Bodley)
- rgw: Swift SLO size\_bytes member is optional ([issue#18936](#), [pr#22967](#), Matt Benjamin)
- rgw: Swift's TempURL can handle temp\_url\_expires written in ISO8601 ([issue#20795](#), [pr#16658](#), Radoslaw Zarzynski)
- rgw: sync module: avoid printing attrs of objects in log ([issue#37646](#), [pr#25541](#),

Abhishek Lekshmanan)

- rgw: test bi list ([issue#24483](#), [pr#21772](#), Orit Wasserman)
- rgw: test/rgw: add ifdef for HAVE\_BOOST\_CONTEXT ([pr#25744](#), Casey Bodley)
- rgw,tests: qa: add test for <https://github.com/ceph/ceph/pull/22790> ([pr#23143](#), yuliyang)
- rgw,tests: qa/rgw: add cls\_lock/log/refcount/version tests to verify suite ([pr#25381](#), Casey Bodley)
- rgw,tests: qa/rgw: add missing import line ([pr#25298](#), Shilpa Jagannath)
- rgw,tests: qa/rgw: add radosgw-admin-rest task to singleton suite ([pr#23145](#), Casey Bodley)
- rgw,tests: qa/rgw: disable testing on ec-cache pools ([issue#23965](#), [pr#22126](#), Casey Bodley)
- rgw,tests: qa/rgw: fix invalid syntax error in radosgw\_admin\_rest.py ([issue#37440](#), [pr#25305](#), Casey Bodley)
- rgw,tests: qa/rgw: move ragweed upgrade test into upgrade/luminous-x ([pr#21707](#), Casey Bodley)
- rgw,tests: qa/rgw: override valgrind -max-threads for radosgw ([issue#25214](#), [pr#23372](#), Casey Bodley)
- rgw,tests: qa/rgw: patch keystone requirements.txt ([issue#23659](#), [pr#23402](#), Casey Bodley)
- rgw,tests: qa/rgw: reduce number of multisite log shards ([pr#24011](#), Casey Bodley)
- rgw,tests: qa/rgw: reorganize verify tasks ([pr#22249](#), Casey Bodley)
- rgw,tests: qa/rgw/tempest: either force os\_type or select random distro ([pr#25996](#), Yehuda Sadeh)
- rgw,tests: test/rgw: fix for bucket checkpoints ([issue#24212](#), [pr#22124](#), Casey Bodley)
- rgw,tests: test/rgw: fix race in test\_rgw\_reshard\_wait ([pr#26741](#), Casey Bodley)
- rgw,tests: test/rgw: silence -Wsign-compare warnings ([pr#26364](#), Kefu Chai)
- rgw: The delete markers generated by object expiration should have owner attribute ([issue#24568](#), [pr#22619](#), Zhang Shaowen)
- rgw: the error code returned by rgw is different from amz s3 when getting cors ([issue#26964](#), [pr#23646](#), ashitakasam)

- rgw: thread DoutPrefixProvider into RGW::Auth\_S3::authorize ([pr#24409](#), Ali Maredia)
- rgw, tools: ceph-dencoder: add RGWRealm and RGWPeriod support ([pr#25057](#), yuliyang)
- rgw, tools: cls: refcount: add obj\_refcount to ceph-dencoder ([pr#25441](#), Abhishek Lekshmanan)
- rgw, tools: cls/rgw: ready rgw\_usage\_log\_entry for extraction via ceph-dencoder ([issue#34537](#), [pr#22344](#), Vaibhav Bhembre)
- rgw, tools: vstart: make beast as the default frontend for rgw ([pr#26566](#), Abhishek Lekshmanan)
- rgw, tools: vstart: rgw: disable the lc debug interval option ([pr#25487](#), Abhishek Lekshmanan)
- rgw, tools: vstart: set admin socket for RGW in conf ([pr#23983](#), Abhishek Lekshmanan)
- rgw: update cls\_rgw.cc and cls\_rgw\_const.h ([pr#24001](#), yuliang)
- rgw: update ObjectCacheInfo::time\_added on overwrite ([issue#24346](#), [pr#22324](#), Casey Bodley)
- rgw: update -url in usage and doc ([pr#22100](#), Jos Collin)
- rgw: use chunked encoding to get partial results out faster ([issue#12713](#), [pr#23940](#), Robin H. Johnson)
- rgw: use coarse\_real\_clock for req\_state::time ([pr#21893](#), Casey Bodley)
- rgw: use DoutPrefixProvider to add more context to log output ([pr#21700](#), Casey Bodley)
- rgw: use partial-order bucket listing in RGWL, add configurable processing delay ([issue#23956](#), [pr#21755](#), Matt Benjamin)
- rgw: User Policy ([pr#21379](#), Pritha Srivastava)
- rgw: user stats account for resharded buckets ([pr#24595](#), Casey Bodley)
- rgw: warn if zone doesn't contain all zg's placement targets ([pr#22452](#), Abhishek Lekshmanan)
- rgw: website routing rules num limit ([pr#23429](#), yuliang)
- rgw: when exclusive lock fails due existing lock, log add'l info ([issue#38171](#), [pr#26272](#), J. Eric Ivancich)
- rgw: zone service only provides const access to its data ([pr#25412](#), Casey Bodley)

- rocksdb: pick up a fix to be backward compatible ([issue#25146](#), [pr#25070](#), Kefu Chai)
- script: build-integration-branch: avoid Unicode error ([issue#24003](#), [pr#21807](#), Nathan Cutler)
- script/kubejacker: Add openSUSE based images ([pr#24055](#), Sebastian Wagner)
- scripts: backport-create-issue: complain about duplicates and support mimic ([issue#24071](#), [pr#21634](#), Nathan Cutler)
- seastar: pickup fix for segfault in POSIX stack ([pr#25861](#), Kefu Chai)
- spec: add missing rbd mirror bootstrap directory ([pr#24856](#), Sébastien Han)
- src: balance std::hex and std::dec manipulators ([pr#22287](#), Kefu Chai)
- src/ceph.in: dev mode: add build path to beginning of PATH, not end ([issue#24578](#), [pr#22628](#), Dan Mick)
- src: Eliminate new warnings in Fedora 28 ([pr#21898](#), Adam C. Emerson)
- test/crimson: fixes of unittest\_seastar\_echo ([pr#26419](#), Yingxin Cheng, Kefu Chai)
- test/fio: fix compiler failure ([pr#22728](#), Jianpeng Ma)
- test/fio: new option to control file preallocation ([pr#23410](#), Igor Fedotov)
- tests: Add hashinfo testing for dump command of ceph-objectstore-tool ([issue#38053](#), [pr#26158](#), David Zafman)
- tests: add ubuntu 18.04 dockerfile ([pr#25251](#), Kefu Chai)
- tests: auth, test: fix building on ARMs after the NSS -> OpenSSL transition ([pr#22129](#), Radoslaw Zarzynski)
- tests: ceph\_kvstorebench: include <errno.h> not asm-generic/errno.h ([pr#25256](#), Kefu Chai)
- tests: ceph-volume: functional tests, add libvirt customization ([pr#25895](#), Jan Fajerski)
- tests: do not check for invalid k/m combinations ([issue#16500](#), [pr#25046](#), Kefu Chai)
- tests: Fixes for standalone tests ([pr#22480](#), David Zafman)
- tests: fix to check server\_conn in MessengerTest.NameAddrTest ([pr#23931](#), Yingxin)
- tests: make ceph-admin-commands.sh log what it does ([issue#37089](#), [pr#25080](#), Nathan Cutler)

- tests: make test\_ceph\_argparse.py pass on py3-only systems ([issue#24816](#), [pr#22922](#), Nathan Cutler)
- tests: mgr/ansible: add install tox==2.9.1 ([pr#26313](#), Kefu Chai)
- tests: mgr/dashboard: Added additional breadcrumb and tab tests to Cluster menu ([pr#26151](#), Nathan Weinberg)
- tests: mgr/dashboard: Added additional breadcrumb tests to Cluster ([pr#25010](#), Nathan Weinberg)
- tests: mgr/dashboard: Added breadcrumb and tab tests to Pools menu ([pr#25572](#), Nathan Weinberg)
- tests: mgr/dashboard: Added breadcrumb tests to Block menu items ([pr#25143](#), Nathan Weinberg)
- tests: mgr/dashboard: Added breadcrumb tests to Filesystems menu ([pr#26592](#), Nathan Weinberg)
- tests: mgr/dashboard: Added NFS Ganesha suite to QA tests ([pr#26510](#), Laura Paduano)
- tests: mgr/dashboard: Added tab tests to Block menu items ([pr#26243](#), Nathan Weinberg)
- tests: mgr/dashboard: Add Jest Runner ([pr#22031](#), Tiago Melo)
- tests: mgr/dashboard: Add unit test case for controller/erasure\_code\_profile.py ([pr#24789](#), Ranjitha G)
- tests: mgr/dashboard: Add unit test for frontend api services ([pr#22284](#), Tiago Melo)
- tests: mgr/dashboard: Add unit tests for all frontend pipes ([pr#22182](#), Tiago Melo)
- tests: mgr/dashboard: Add unit test to the frontend services ([pr#22244](#), Tiago Melo)
- tests: mgr/dashboard: Fix a broken ECP controller test ([pr#25363](#), Zack Cerza)
- tests: mgr/dashboard: Fix PYTHONPATH for test runner ([pr#25359](#), Zack Cerza)
- tests: mgr/dashboard: Improve max-line-length tslint rule ([pr#22279](#), Tiago Melo)
- tests: mgr/dashboard: RbdMirroringService test suite fails in dev mode ([issue#37841](#), [pr#25865](#), Stephan Müller)
- tests: mgr/dashboard: Small improvements for running teuthology tests ([pr#25121](#), Zack Cerza)

- tests: mgr/dashboard: updated API test ([pr#25653](#), Alfonso Martínez)
- tests: mgr/dashboard: updated API test to reflect changes in ModuleInfo ([pr#25761](#), Kefu Chai)
- tests: mgr/test\_orchestrator: correct ceph-volume path ([issue#37773](#), [pr#25839](#), Kefu Chai)
- tests: object errors found in be\_select\_auth\_object() aren't logged the same ([issue#25108](#), [pr#23376](#), David Zafman)
- tests: osd/OSDMap: set pg\_autoscale\_mode with setting from conf ([pr#25746](#), Kefu Chai)
- tests: os/tests: fix garbageCollection test case from store\_test suite ([pr#23752](#), Igor Fedotov)
- tests: os/tests: silence -Wsign-compare warning ([pr#25072](#), Kefu Chai)
- tests: qa: add librados3 to exclude\_packages for upgrade tests ([pr#25037](#), Kefu Chai)
- tests: qa: add test that builds example librados programs ([issue#35989](#), [issue#15100](#), [pr#23131](#), Nathan Cutler)
- tests: qa/ceph-ansible: Set ceph\_stable\_release to mimic ([issue#38231](#), [pr#26328](#), Brad Hubbard)
- tests: qa/distros: add openSUSE Leap 42.3 and 15.0 ([pr#24380](#), Nathan Cutler)
- tests: qa: Don't use sudo when moving logs ([pr#22763](#), David Zafman)
- tests: qa: downgrade librados2,librbd1 for thrash-old-clients tests ([issue#37618](#), [pr#25463](#), Kefu Chai)
- tests: qa: fix manager module paths ([pr#23637](#), Noah Watkins, David Zafman)
- tests/qa - fix mimic subset for nightlies ([pr#21931](#), Yuri Weinstein)
- tests: qa: fix test on "ceph fs set cephfs allow\_new\_snaps" ([pr#21829](#), Kefu Chai)
- tests: qa: fix upgrade tests and test\_envlibrados\_for\_rocksdb.sh ([pr#25106](#), Kefu Chai)
- tests: qa: For teuthology copy logs to teuthology expected location ([pr#22702](#), David Zafman)
- tests: qa/mgr/dashboard: Fix type annotation error ([pr#25235](#), Sebastian Wagner)
- tests: qa/mon: fix cluster support for monmap bootstrap ([issue#38115](#), [pr#26205](#), Casey Bodley)

- tests: qa/standalone: Minor test improvements ([issue#35912](#), [pr#24018](#), David Zafman)
- tests: qa/standalone/scrub: When possible show side-by-side diff in addition to regular diff ([pr#22727](#), David Zafman)
- tests: qa/standalone: Standalone test corrections ([issue#35982](#), [pr#24088](#), David Zafman)
- tests: qa/suites/rados/upgrade: remove stray link ([pr#22460](#), Sage Weil)
- tests: qa/suites/rados/upgrade: set require-osd-release to nautilus ([issue#37432](#), [pr#25314](#), Kefu Chai)
- tests: qa/suites/rados/verify: remove random-distro\$ ([pr#22057](#), Kefu Chai)
- tests: qa/suites/upgrade/mimic-x: fix rhel runs ([pr#25781](#), Neha Ojha)
- tests: qa/tasks/mgr: fix test\_pool.py ([issue#24077](#), [pr#21943](#), Kefu Chai)
- tests: qa/tasks/thrashosds-health.yaml: whitelist slow requests ([issue#25104](#), [pr#23237](#), Neha Ojha)
- tests: qa/tasks: update mirror link for maven ([pr#23944](#), Vasu Kulkarni)
- tests: qa/tests: added filters to support distro tests for client-upgrade tests ([pr#22096](#), Yuri Weinstein)
- tests: qa/tests - added mimic-p2p suite ([pr#22726](#), Yuri Weinstein)
- tests: qa/tests: Added mimic runs, removed large suites (rados, rbd, etc) ru... ([pr#21827](#), Yuri Weinstein)
- tests: qa/tests: added “-n 7” to make sure mimic-x runs on built master branch ([pr#25038](#), Yuri Weinstein)
- tests: qa/tests: added rhel 7.6 ([pr#25919](#), Yuri Weinstein)
- tests: qa/tests: fix volume size when running in ovh ([pr#21961](#), Vasu Kulkarni)
- tests: qa/tests: Move ceph-ansible tests to ansible version 2.7 ([issue#37973](#), [pr#26068](#), Brad Hubbard)
- tests: qa/tests: remove ceph-disk tests from ceph-deploy and default all tests to use ceph-volume ([pr#22921](#), Vasu Kulkarni)
- tests: qa/upgrade: cleanup for nautilus ([pr#23305](#), Nathan Cutler)
- tests: qa: use \$TESTDIR for testing mkfs ([pr#22246](#), Kefu Chai)
- tests: qa: wait longer for osd to flush pg stats ([issue#24321](#), [pr#22275](#), Kefu Chai)

- tests: qa/workunits/ceph-disk: -no-mon-config ([pr#21942](#), Kefu Chai)
- tests: qa/workunits/mon/test\_mon\_config\_key.py: bump up the size limit ([issue#36260](#), [pr#24340](#), Kefu Chai)
- tests: qa/workunits/rados/test\_envlibrados\_for\_rocksdb: install g++ not g++-4.7 ([pr#22103](#), Kefu Chai)
- tests: qa/workunits/rados/test\_librados\_build.sh: grab files from explicit git branch ([pr#25268](#), Nathan Cutler)
- tests: run-make-check: increase fs.aio-max-nr to 1048576 ([pr#23689](#), Kefu Chai)
- tests: test/common: silence GCC warnings ([pr#23692](#), Kefu Chai)
- tests: test/crimson: add dummy\_auth to test\_async\_echo ([pr#26783](#), Yingxin Cheng)
- tests: test/crimson: fix build failure of test\_alien\_echo ([pr#26308](#), chunmei Liu)
- tests: test/crimson: fix FTBFS of unittest\_seastar\_perfcounters on arm64 ([pr#25647](#), Kefu Chai)
- tests: test/crimson: split async-msgr out of alien\_echo ([pr#26620](#), Yingxin Cheng)
- tests: test/dashboard: fix segfault when importing dm.xmlsec.binding ([issue#37081](#), [pr#25139](#), Kefu Chai)
- tests: test: Disable duplicate request command test during scrub testing ([pr#25675](#), David Zafman)
- tests: test/docker-test-helper.sh: move “cp .git/HEAD” out of loop ([pr#22978](#), Kefu Chai)
- tests: test/encoding: Fix typo in encoding/types.h file ([pr#22332](#), TommyLike)
- tests: test/fio: pass config params to object store in a different manner ([pr#23267](#), Igor Fedotov)
- tests: test: fix compile error in test/crimson/test\_config.cc ([pr#23724](#), Yingxin)
- tests: test: fix libc++ crash in Log.GarbleRecovery ([pr#25135](#), Casey Bodley)
- tests: test/librados: fix LibRadosList.ListObjectsNS ([pr#22771](#), Kefu Chai)
- tests: test: Limit loops waiting for force-backfill/force-recovery to happen ([issue#38309](#), [pr#26416](#), David Zafman)
- tests: test: Need to escape parens in log-whitelist for grep ([pr#22074](#), David Zafman)
- tests: test: osd-backfill-stats.sh Fix check of multi backfill OSDs, skip re... ([pr#26330](#), David Zafman)

- tests: test/pybind/test\_rados.py: collect output in stdout for “bench” cmd ([pr#21957](#), Kefu Chai)
- tests: test: run-standalone.sh: point LD\_LIBRARY\_PATH to \$(pwd)/lib ([issue#38262](#), [pr#26371](#), David Zafman)
- tests: tests/qa: trying \$ distro mix ([pr#21895](#), Yuri Weinstein)
- tests: test: Start using GNU awk and fix archiving directory ([pr#23955](#), Willem Jan Withagen)
- tests: test strtol: add test case for parsing hex numbers ([pr#21582](#), Jan Fajerski)
- tests: test: suppress core dumping in there tests as well ([pr#25311](#), Willem Jan Withagen)
- tests: test: switch to GNU sed on FreeBSD ([pr#26318](#), Willem Jan Withagen)
- tests: test: test\_get\_timeout\_delays() fix ([pr#22837](#), David Zafman)
- tests: test: Use a file that should be on all OSes ([pr#22428](#), David Zafman)
- tests: test: Use a grep pattern that works across releases ([issue#35845](#), [pr#24013](#), David Zafman)
- tests: test: Use pids instead of jobspecs which were wrong ([issue#27056](#), [pr#23695](#), David Zafman)
- tests: test: wait\_for\_pg\_stats() should do another check after last 13 secon... ([pr#22198](#), David Zafman)
- tests: test: Whitelist corrections ([pr#22164](#), David Zafman)
- tests: test: write log file to current directory ([issue#36737](#), [pr#25704](#), Kefu Chai)
- tests,tools: ceph-objectstore-tool: Dump hashinfo ([issue#37597](#), [pr#25483](#), David Zafman)
- tests: update Dockerfile to support fc-29 ([pr#26311](#), Kefu Chai)
- tests: upgrade/luminous-x: fix order of final-workload directory ([pr#23162](#), Nathan Cutler)
- tests: upgrade/luminous-x: whitelist REQUEST\_SLOW for rados\_mon\_thrash ([issue#25051](#), [pr#23160](#), Nathan Cutler)
- tests: Wip 38027 38195: osd/osd-backfill-space.sh fails ([issue#38027](#), [issue#38195](#), [pr#26290](#), David Zafman)
- tools: Add clear-data-digest command to objectstore tool ([pr#25403](#), Li Yichao)

- tools: add offset-align option to “rados” load-gen ([pr#20683](#), Zengran Zhang)
- tools: backport-create-issue: rate-limit to avoid seeming like a spammer ([pr#24243](#), Nathan Cutler)
- tools: ceph-menv: mrun shell environment ([pr#22132](#), Yehuda Sadeh)
- tools: ceph-objectstore-tool: Allow target level as first positional argument ([issue#35846](#), [pr#23989](#), David Zafman)
- tools: correct the description of Allowed options in osdomap tool ([pr#23488](#), xiaomanh)
- tools, mgr: silence clang warnings ([pr#23430](#), Kefu Chai)
- tools: mstop.sh allow kill -9 after failing to kill procs ([pr#26680](#), Yuval Lifshitz)
- tools/rados: fix memory leak in error path ([pr#25410](#), Li Wang)
- tools: script/kubejacker: include cls libs ([pr#23569](#), John Spray)
- tools: script: new ceph-backport.sh script ([pr#22875](#), Nathan Cutler)
- tools: tools: ceph-authtool: report correct number of caps when creating keyring ([pr#23304](#), Nathan Cutler)
- tools: tools/ceph\_kvstore\_tool: do not open rocksdb when repairing it ([pr#25108](#), Kefu Chai)
- tools: tools/ceph\_kvstore\_tool: extract StoreTool into kvstore\_tool.cc ([pr#26041](#), Kefu Chai)
- tools: tools/ceph\_kvstore\_tool: Move summary output to print\_summary ([pr#26666](#), Brad Hubbard)
- tools: tools/rados: allow list objects in a specific pg in a pool ([pr#19041](#), Li Wang)
- tools: tools/rados: always call rados.shutdown() before exit() ([issue#36732](#), [pr#24990](#), Li Wang)
- tools: tools/rados: correct the read offset of bench ([pr#23667](#), Xiaofei Cui)
- tools: tools/rados: fix the unit of target-throughput ([pr#23683](#), Xiaofei Cui)
- vstart: disable dashboard when rbd not built ([pr#23336](#), Noah Watkins)
- vstart.sh: fix params generation for monmaptool ([issue#38174](#), [pr#26273](#), Yehuda Sadeh)

# Archived Releases

---

- [v13.2.10 Mimic](#)
- [v13.2.9 Mimic](#)
- [v13.2.8 Mimic](#)
- [v13.2.7 Mimic](#)
- [v13.2.6 Mimic](#)
- [v13.2.5 Mimic](#)
- [v13.2.4 Mimic](#)
- [v13.2.3 Mimic](#)
- [v13.2.2 Mimic](#)
- [v13.2.1 Mimic](#)
- [v13.2.0 Mimic](#)
- [v12.2.13 Luminous](#)
- [v12.2.12 Luminous](#)
- [v12.2.11 Luminous](#)
- [v12.2.10 Luminous](#)
- [v12.2.9 Luminous](#)
- [v12.2.8 Luminous](#)
- [v12.2.7 Luminous](#)
- [v12.2.6 Luminous](#)
- [v12.2.5 Luminous](#)
- [v12.2.4 Luminous](#)
- [v12.2.3 Luminous](#)
- [v12.2.2 Luminous](#)
- [v12.2.1 Luminous](#)
- [v12.2.0 Luminous](#)
- [v11.2.1 Kraken](#)
- [v11.2.0 Kraken](#)
- [v11.0.2 Kraken](#)
- [v10.2.11 Jewel](#)
- [v10.2.10 Jewel](#)
- [v10.2.9 Jewel](#)
- [v10.2.8 Jewel](#)
- [v10.2.7 Jewel](#)
- [v10.2.6 Jewel](#)
- [v10.2.5 Jewel](#)
- [v10.2.4 Jewel](#)
- [v10.2.3 Jewel](#)
- [v10.2.2 Jewel](#)
- [v10.2.1 Jewel](#)
- [v10.2.0 Jewel](#)
- [v9.2.1 Infernalis](#)
- [v9.2.0 Infernalis](#)

- [v9.1.0 Infernalis release candidate](#)
- [v9.0.3](#)
- [v9.0.2](#)
- [v9.0.1](#)
- [v9.0.0](#)
- [v0.94.10 Hammer](#)
- [v0.94.9 Hammer](#)
- [v0.94.8 Hammer](#)
- [v0.94.7 Hammer](#)
- [v0.94.6 Hammer](#)
- [v0.94.5 Hammer](#)
- [v0.94.4 Hammer](#)
- [v0.94.3 Hammer](#)
- [v0.94.2 Hammer](#)
- [v0.94.1 Hammer](#)
- [v0.94 Hammer](#)
- [v0.93](#)
- [v0.92](#)
- [v0.91](#)
- [v0.90](#)
- [v0.89](#)
- [v0.88](#)
- [v0.87.2 Giant](#)
- [v0.87.1 Giant](#)
- [v0.87 Giant](#)
- [v0.86](#)
- [v0.85](#)
- [v0.84](#)
- [v0.83](#)
- [v0.82](#)
- [v0.81](#)
- [v0.80.11 Firefly](#)
- [v0.80.10 Firefly](#)
- [v0.80.9 Firefly](#)
- [v0.80.8 Firefly](#)
- [v0.80.7 Firefly](#)
- [v0.80.6 Firefly](#)
- [v0.80.5 Firefly](#)
- [v0.80.4 Firefly](#)
- [v0.80.3 Firefly](#)
- [v0.80.2 Firefly](#)
- [v0.80.1 Firefly](#)
- [v0.80 Firefly](#)
- [v0.79](#)
- [v0.78](#)
- [v0.77](#)

- [v0.76](#)
- [v0.75](#)
- [v0.74](#)
- [v0.73](#)
- [v0.72.3 Emperor \(pending release\)](#)
- [v0.72.2 Emperor](#)
- [v0.72.1 Emperor](#)
- [v0.72 Emperor](#)
- [v0.71](#)
- [v0.70](#)
- [v0.69](#)
- [v0.68](#)
- [v0.67.12 "Dumpling" \(draft\)](#)
- [v0.67.11 "Dumpling"](#)
- [v0.67.10 "Dumpling"](#)
- [v0.67.9 "Dumpling"](#)
- [v0.67.8 "Dumpling"](#)
- [v0.67.7 "Dumpling"](#)
- [v0.67.6 "Dumpling"](#)
- [v0.67.5 "Dumpling"](#)
- [v0.67.4 "Dumpling"](#)
- [v0.67.3 "Dumpling"](#)
- [v0.67.2 "Dumpling"](#)
- [v0.67.1 "Dumpling"](#)
- [v0.67 "Dumpling"](#)
- [v0.66](#)
- [v0.65](#)
- [v0.64](#)
- [v0.63](#)
- [v0.62](#)
- [v0.61.9 "Cuttlefish"](#)
- [v0.61.8 "Cuttlefish"](#)
- [v0.61.7 "Cuttlefish"](#)
- [v0.61.6 "Cuttlefish"](#)
- [v0.61.5 "Cuttlefish"](#)
- [v0.61.4 "Cuttlefish"](#)
- [v0.61.3 "Cuttlefish"](#)
- [v0.61.2 "Cuttlefish"](#)
- [v0.61.1 "Cuttlefish"](#)
- [v0.61 "Cuttlefish"](#)
- [v0.60](#)
- [v0.59](#)
- [v0.58](#)
- [v0.57](#)
- [v0.56.7 "bobtail"](#)
- [v0.56.6 "bobtail"](#)

- [v0.56.5 “bobtail”](#)
- [v0.56.4 “bobtail”](#)
- [v0.56.3 “bobtail”](#)
- [v0.56.2 “bobtail”](#)
- [v0.56.1 “bobtail”](#)
- [v0.56 “bobtail”](#)
- [v0.54](#)
- [v0.48.3 “argonaut”](#)
- [v0.48.2 “argonaut”](#)
- [v0.48.1 “argonaut”](#)
- [v0.48 “argonaut”](#)

## v13.2.10 Mimic

This is the tenth bugfix release of Ceph Mimic, this release fixes a RGW vulnerability affecting `mimic`, and we recommend that all `mimic` users upgrade.

## Notable Changes

- CVE 2020 12059: Fix an issue with Post Object Requests with Tagging ([issue#44967](#), Lei Cao, Abhishek Lekshmanan)

## v13.2.9 Mimic

This is the ninth and very likely the last stable release in the Ceph Mimic stable release series. This release fixes bugs across all components and also contains a RGW security fix. We recommend all `mimic` users to upgrade to this version.

## Notable Changes

- CVE-2020-1760: Fixed XSS due to RGW GetObject header-splitting
- The configuration value `osd_calc_pg_upmaps_max_stddev` used for upmap balancing has been removed. Instead use the mgr balancer config `upmap_max_deviation` which now is an integer number of PGs of deviation from the target PGs per OSD. This can be set with a command like `ceph config set mgr mgr/balancer/upmap_max_deviation 2`. The default `upmap_max_deviation` is 1. There are situations where crush rules would not allow a pool to ever have completely balanced PGs. For example, if crush requires 1 replica on each of 3 racks, but there are fewer OSDs in 1 of the racks. In those cases, the configuration value can be increased.
- The `cephfs-data-scan scan_links` command now automatically repair inotables and snaptable.

## Changelog

- bluestore: os/bluestore: fix improper setting of STATE\_KV\_SUBMITTED ([pr#31673](#), Igor Fedotov)
- ceph-volume/batch: check lvs list before access ([pr#34479](#), Jan Fajerski)
- ceph-volume/batch: fail on filtered devices when non-interactive ([pr#33201](#), Jan Fajerski)
- ceph-volume/batch: return success when all devices are filtered ([pr#34476](#), Jan Fajerski)

Fajerski)

- ceph-volume/lvm/activate.py: clarify error message: fsid refers to osd\_fsid ([pr#32865](#), Yaniv Kaul)
- ceph-volume/test: patch VolumeGroups ([pr#32559](#), Jan Fajerski)
- ceph-volume: Dereference symlink in lvm list ([pr#32876](#), Benoît Knecht)
- ceph-volume: add db and wal support to raw mode ([pr#33622](#), Sébastien Han)
- ceph-volume: add methods to pass filters to pvs, vgs and lvs commands ([pr#33215](#), Rishabh Dave)
- ceph-volume: add proper size attribute to partitions ([pr#32529](#), Jan Fajerski)
- ceph-volume: add raw mode ([pr#33580](#), Jan Fajerski, Sage Weil, Guillaume Abrioux)
- ceph-volume: add sizing arguments to prepare ([pr#33578](#), Jan Fajerski)
- ceph-volume: add utility functions ([pr#32544](#), Mohamad Gebai)
- ceph-volume: allow raw block devices everywhere ([pr#32869](#), Jan Fajerski)
- ceph-volume: allow to skip restorecon calls ([pr#32530](#), Alfredo Deza)
- ceph-volume: avoid calling zap\_lv with a LV-less VG ([pr#33610](#), Jan Fajerski)
- ceph-volume: batch bluestore fix create\_lvs call ([pr#33579](#), Jan Fajerski)
- ceph-volume: batch bluestore fix create\_lvs call ([pr#33623](#), Jan Fajerski)
- ceph-volume: check if we run in an selinux environment ([pr#32866](#), Jan Fajerski)
- ceph-volume: check if we run in an selinux environment, now also in py2 ([pr#32867](#), Jan Fajerski)
- ceph-volume: devices/simple/scan: Fix string in log statement ([pr#34444](#), Jan Fajerski)
- ceph-volume: don't create osd['block.db'] by default ([pr#33626](#), Jan Fajerski)
- ceph-volume: don't remove vg twice when zapping filestore ([pr#33615](#), Jan Fajerski)
- ceph-volume: finer grained availability notion in inventory ([pr#33606](#), Jan Fajerski)
- ceph-volume: fix is\_ceph\_device for lvm batch ([pr#33608](#), Jan Fajerski, Dimitri Savineau)
- ceph-volume: fix the integer overflow ([pr#32872](#), dongdong tao)

- ceph-volume: import mock.mock instead of unittest.mock (py2) ([pr#32871](#), Jan Fajerski)
- ceph-volume: lvm deactivate command ([pr#33208](#), Jan Fajerski)
- ceph-volume: lvm/deactivate: add unit tests, remove -all ([pr#32862](#), Jan Fajerski)
- ceph-volume: lvm: get\_device\_vgs() filter by provided prefix ([pr#33617](#), Jan Fajerski, Yehuda Sadeh)
- ceph-volume: make get\_devices fs location independent ([pr#33124](#), Jan Fajerski)
- ceph-volume: minor clean-up of “simple scan” subcommand help ([pr#32557](#), Michael Fritch)
- ceph-volume: mokeypatch calls to lvm related binaries ([pr#31406](#), Jan Fajerski)
- ceph-volume: pass journal\_size as Size not string ([pr#33611](#), Jan Fajerski)
- ceph-volume: rearrange api/lvm.py ([pr#31407](#), Rishabh Dave)
- ceph-volume: refactor listing.py + fixes ([pr#33603](#), Jan Fajerski, Rishabh Dave, Theofilos Mouratidis, Guillaume Abrioux)
- ceph-volume: reject disks smaller then 5GB in inventory ([issue#40776](#), [pr#32528](#), Jan Fajerski)
- ceph-volume: silence ‘ceph-bluestore-tool’ failures ([pr#33605](#), Sébastien Han)
- ceph-volume: skip missing interpreters when running tox tests ([pr#33489](#), Andrew Schoen)
- ceph-volume: skip osd creation when already done ([pr#33607](#), Guillaume Abrioux)
- ceph-volume: strip \_dmcrypt suffix in simple scan json output ([pr#33618](#), Jan Fajerski)
- ceph-volume: use correct extents if using db-devices and >1 osds\_per\_device ([pr#32875](#), Fabian Niepelt)
- ceph-volume: use fsync for dd command ([pr#31552](#), Rishabh Dave)
- ceph-volume: use get\_device\_vgs in has\_common\_vg ([pr#33609](#), Jan Fajerski)
- ceph-volume: util: look for executable in \$PATH ([pr#32861](#), Shyukri Shyukriev)
- cephfs: cephfs: osdc/objecter: Fix last\_sent in scientific format and add age to ops ([pr#31384](#), Varsha Rao)
- cephfs: cephfs: test\_volume\_client: declare only one default for python version ([issue#40460](#), [pr#30110](#), Rishabh Dave)

- cephfs: client: more precise CEPH\_CLIENT\_CAPS\_PENDING\_CAPSNAP ([pr#31283](#), “Yan, Zheng”)
- cephfs: client: remove Iinode.dir\_contacts field and handle bad whence value to llseek gracefully ([pr#31380](#), Jeff Layton)
- cephfs: mds: avoid calling clientreplay\_done() prematurely ([pr#31282](#), “Yan, Zheng”)
- cephfs: mds: fix assert(omap\_num\_objs <= MAX\_OBJECTS) of OpenFileTable ([pr#32757](#), “Yan, Zheng”)
- cephfs: mds: fix infinite loop in Locker::file\_update\_finish ([pr#31284](#), “Yan, Zheng”)
- cephfs: mds: mds returns -5(EIO) error when the deleted file does not exist ([pr#31381](#), huanwen ren)
- cephfs: mds: split the dir if the op makes it oversized, because some ops maybe in flight ([pr#31379](#), simon gao)
- cephfs: tools/cephfs: make ‘cephfs-data-scan scan\_links’ reconstruct snaptable ([pr#31281](#), “Yan, Zheng”)
- common/config: parse -log-early option ([pr#33130](#), Sage Weil)
- common: common/admin\_socket: Increase socket timeouts ([pr#33323](#), Brad Hubbard)
- common: common/config: update values when they are removed via mon ([pr#33327](#), Sage Weil)
- common: common/util: use ifstream to read from /proc files ([pr#32902](#), Kefu Chai, songweibin)
- core,mgr,tests: mgr: Release GIL and Balancer fixes ([pr#31957](#), Neha Ojha, Kefu Chai, Noah Watkins, David Zafman)
- core,mgr: mgr/prometheus: assign a value to osd\_dev\_node when obj\_store is not filestore or bluestore ([pr#31557](#), jiahui.zeng)
- core,tests: qa/tasks/cbt: install python3 deps ([pr#34193](#), Sage Weil)
- core: mon/OSDMonitor: fix format error ceph osd stat -format json ([pr#33322](#), Zheng Yin)
- core: mon: Don’t put session during feature change ([pr#33154](#), Brad Hubbard)
- core: osd/PeeringState.cc: don’t let num\_objects become negative ([pr#33331](#), Neha Ojha)
- core: osd/PeeringState.cc: skip peer\_purged when discovering all missing ([pr#33329](#), Neha Ojha)

- core: osd/PeeringState.h: ignore MLogRec in Peering/GetInfo ([pr#33594](#), Neha Ojha)
- core: osd/PeeringState: do not exclude up from acting\_recovery\_backfill ([pr#33324](#), Nathan Cutler, xie xingguo)
- core: osd: Allow 64-char hostname to be added as the “host” in CRUSH ([pr#33145](#), Michal Skalski)
- core: osd: Diagnostic logging for upmap cleaning ([pr#32717](#), David Zafman)
- core: osd: backfill\_toofull seen on cluster where the most full OSD is at 1% ([pr#32361](#), David Zafman)
- core: osd: set collection pool opts on collection create, pg load ([pr#32125](#), Sage Weil)
- core: selinux: Allow ceph to read udev db ([pr#32258](#), Boris Ranto)
- core: selinux: Allow ceph-mgr access to httpd dir ([pr#34458](#), Brad Hubbard)
- doc: remove invalid option mon\_pg\_warn\_max\_per\_osd ([pr#31875](#), zhang daolong)
- doc: doc/\_templates/page.html: redirect to etherpad ([pr#32249](#), Neha Ojha)
- doc: doc/cephfs/client-auth: description and example are inconsistent ([pr#32782](#), Ilya Dryomov)
- doc: wrong datatype describing crush\_rule ([pr#32255](#), Kefu Chai)
- mgr,pybind: mgr/prometheus: report per-pool pg states ([pr#33158](#), Aleksei Zakharov)
- mgr,pybind: mgr/telemetry: check get\_metadata return val ([pr#33096](#), Yaarit Hatuka)
- mount.ceph: give a hint message when no mds is up or cluster is laggy ([pr#32911](#), Xiubo Li)
- pybind: pybind/mgr: Cancel output color control ([pr#31805](#), Zheng Yin)
- qa: get rid of iterkeys for py3 compatibility ([pr#33999](#), Kyr Shatskyy)
- rbd: creating thick-provision image progress percent info exceeds 100% ([pr#33318](#), Xiangdong Mu)
- rbd: librbd: diff iterate with fast-diff now correctly includes parent ([pr#32470](#), Jason Dillaman)
- rbd: librbd: don't call refresh from mirror::GetInfoRequest state machine ([pr#32952](#), Mykola Golub)
- rbd: librbd: fix rbd\_open\_by\_id, rbd\_open\_by\_id\_read\_only ([pr#33315](#), yangjun)

- rbd: nautilus: rbd-mirror: fix ‘rbd mirror status’ asok command output ([pr#32714](#), Mykola Golub)
- rbd: rbd-mirror: clone v2 mirroring improvements ([pr#31520](#), Mykola Golub)
- rbd: rbd-mirror: improve detection of blacklisted state ([pr#33598](#), Mykola Golub)
- rbd: rbd-mirror: make logrotate work ([pr#32598](#), Mykola Golub)
- rgw: add bucket permission verify when copy obj ([pr#31377](#), NancySu05)
- rgw: add list user admin OP API ([pr#31754](#), Oshyn Song)
- rgw: add missing admin property when sync user info ([pr#30804](#), zhang Shaowen)
- rgw: add num\_shards to radosgw-admin bucket stats ([pr#31183](#), Paul Emmerich)
- rgw: adding mfa code validation when bucket versioning status is changed ([pr#33303](#), Pritha Srivastava)
- rgw: allow reshards log entries for non-existent buckets to be cancelled ([pr#33302](#), J. Eric Ivancich)
- rgw: auto-clean reshards queue entries for non-existent buckets ([pr#33300](#), J. Eric Ivancich)
- rgw: change the “rgw admin status” ‘num\_shards’ output to signed int ([issue#37645](#), [pr#33305](#), Mark Kogan)
- rgw: crypt: permit RGW-AUTO/default with SSE-S3 headers ([pr#31861](#), Matt Benjamin)
- rgw: find oldest period and update RGWMetadataLogHistory() ([pr#33309](#), Shilpa Jagannath)
- rgw: fix a bug that bucket instance obj can’t be removed after resharding completed ([pr#33306](#), zhang Shaowen)
- rgw: fix bad user stats on versioned bucket after reshards ([pr#33304](#), J. Eric Ivancich)
- rgw: fix memory growth while deleting objects with ([pr#31378](#), Mark Kogan)
- rgw: get barbican secret key request maybe return error code ([pr#33966](#), Richard Bai(白学余))
- rgw: make max\_connections configurable in beast ([pr#33341](#), Tiago Pasqualini)
- rgw: making implicit\_tenants backwards compatible ([issue#24348](#), [pr#33748](#), Marcus Watts)
- rgw: maybe core dump when reload operator happened ([pr#33313](#), Richard Bai(白学余))

- rgw: move forward marker even in case of many rgw.none indexes ([pr#33311](#), Ilsoo Byun)
- rgw: prevent bucket reshards scheduling if bucket is resharding ([pr#31299](#), J. Eric Ivancich)
- rgw: update the hash source for multipart entries during resharding ([pr#33312](#), dongdong tao)

## v13.2.8 Mimic

---

This is the eighth release in the Ceph Mimic stable release series. Its sole purpose is to fix a regression that found its way into the previous release.

### Notable Changes

- Due to a missed backport, clusters in the process of being upgraded from 13.2.6 to 13.2.7 might suffer an OSD crash in `build_incremental_map_msg`. This regression was reported in <https://tracker.ceph.com/issues/43106> and is fixed in 13.2.8 (this release). Users of 13.2.6 can upgrade to 13.2.8 directly - i.e., skip 13.2.7 - to avoid this.

### Changelog

- osd: fix sending incremental map messages (more) ([issue#43106](#), [pr#32000](#), Sage Weil)
- tests: added missing point release versions ([pr#32087](#), Yuri Weinstein)
- tests: rgw: add missing force-branch: ceph-mimic for swift tasks ([pr#32033](#), Casey Bodley)

## v13.2.7 Mimic

---

This is the seventh bugfix release of the Mimic v13.2.x long-term stable release series. All Mimic users are advised to upgrade.

### Notable Changes

MDS:

- Cache trimming is now throttled. Dropping the MDS cache via the “ceph tell mds <foo> cache drop” command or large reductions in the cache size will no longer cause service unavailability.

- Behavior with recalling caps has been significantly improved to not attempt recalling too many caps at once, leading to instability. MDS with a large cache (64GB+) should be more stable.
- MDS now provides a config option “`mds_max_caps_per_client`” (default: 1M) to limit the number of caps a client session may hold. Long running client sessions with a large number of caps have been a source of instability in the MDS when all of these caps need to be processed during certain session events. It is recommended to not unnecessarily increase this value.
- The “`mds_recall_state_timeout`” config parameter has been removed. Late client recall warnings are now generated based on the number of caps the MDS has recalled which have not been released. The new config parameters “`mds_recall_warning_threshold`” (default: 32K) and “`mds_recall_warning_decay_rate`” (default: 60s) set the threshold for this warning.
- The “cache drop” admin socket command has been removed. The “`ceph tell mds.X cache drop`” remains.

#### OSD:

- A health warning is now generated if the average osd heartbeat ping time exceeds a configurable threshold for any of the intervals computed. The OSD computes 1 minute, 5 minute and 15 minute intervals with average, minimum and maximum values. New configuration option “`mon_warn_on_slow_ping_ratio`” specifies a percentage of “`osd_heartbeat_grace`” to determine the threshold. A value of zero disables the warning. A new configuration option “`mon_warn_on_slow_ping_time`”, specified in milliseconds, overrides the computed value, causing a warning when OSD heartbeat pings take longer than the specified amount. A new admin command “`ceph daemon mgr.# dump_osd_network [threshold]`” lists all connections with a ping time longer than the specified threshold or value determined by the config options, for the average for any of the 3 intervals. A new admin command “`ceph daemon osd.# dump_osd_network [threshold]`” does the same but only including heartbeats initiated by the specified OSD.
- The default value of the “`osd_deep_scrub_large_omap_object_key_threshold`” parameter has been lowered to detect an object with large number of omap keys more easily.

#### RGW:

- `radosgw-admin` introduces two subcommands that allow the managing of expire-stale objects that might be left behind after a bucket reshards in earlier versions of RGW. One subcommand lists such objects and the other deletes them. Read the troubleshooting section of the dynamic resharding docs for details.

## Changelog

---

- bluestore: 50-100% iops lost due to bluefs\_preextend\_wal\_files = false ([issue#40280](#), [pr#28574](#), Vitaliy Filippov)
- bluestore: Change default for bluestore\_fsck\_on\_mount\_deep as false ([pr#29699](#), Neha Ojha)
- bluestore: \_txc\_add\_transaction error (39) Directory not empty not handled on operation 21 (op 1, counting from 0) ([issue#39692](#), [pr#29217](#), Sage Weil)
- bluestore: apply shared\_alloc\_size to shared device with log level change ([pr#30219](#), Vikhyat Umrao, Josh Durgin, Sage Weil, Igor Fedotov)
- bluestore: avoid length overflow in extents returned by Stupid Allocator ([issue#40758](#), [issue#40703](#), [pr#29024](#), Igor Fedotov)
- bluestore: common/options: Set concurrent bluestore rocksdb compactions to 2 ([pr#30150](#), Mark Nelson)
- bluestore: default to bitmap allocator for bluestore/bluefs ([pr#28970](#), Igor Fedotov)
- bluestore: dump before “no-spanning blob id” abort ([pr#28029](#), Igor Fedotov)
- bluestore: fix >2GB bluefs writes ([pr#28967](#), Sage Weil, kungf)
- bluestore: fix duplicate allocations in bmap allocator ([issue#40080](#), [pr#28645](#), Igor Fedotov)
- bluestore: load OSD all compression settings unconditionally ([issue#40480](#), [pr#28894](#), Igor Fedotov)
- build/ops: Cython 0.29 removed support for subinterpreters: raises ImportError: Interpreter change detected ([issue#39593](#), [issue#39592](#), [pr#27971](#), Kefu Chai, Tim Serong)
- build/ops: admin/build-doc: use python3 ([pr#30663](#), Kefu Chai)
- build/ops: admin/build-doc: use python3 (follow-on fix) ([pr#30687](#), Nathan Cutler)
- build/ops: backport miscellaneous install-deps.sh and ceph.spec.in fixes from master ([issue#37707](#), [pr#30718](#), Jeff Layton, Kefu Chai, Nathan Cutler, Brad Hubbard, Changcheng Liu, Sebastian Wagner, Yunchuan Wen, Tomasz Setkowski, Zack Cerza)
- build/ops: ceph.spec.in: reserve 2500MB per build job ([pr#30355](#), Dan van der Ster)
- build/ops: cmake,run-make-check.sh: disable SPDK by default ([pr#30183](#), Kefu Chai)
- build/ops: cmake: detect armv8 crc and crypto feature using CHECK\_C\_COMPILER\_FLAG ([issue#17516](#), [pr#30713](#), Kefu Chai)

- build/ops: do\_cmake.sh: source not found ([issue#39981](#), [issue#40005](#), [pr#28217](#), Nathan Cutler)
- build/ops: fix build fail related to PYTHON\_EXECUTABLE variable ([pr#30260](#), Ilsoo Byun)
- build/ops: install-deps.sh: Remove CR repo ([issue#13997](#), [pr#30128](#), Alfredo Deza, Brad Hubbard)
- build/ops: install-deps.sh: install python\*-devel for python\*rpm-macros ([pr#30244](#), Kefu Chai)
- build/ops: make “patch” build dependency explicit ([issue#40269](#), [issue#40175](#), [pr#29150](#), Nathan Cutler)
- build/ops: python3-cephfs should provide python36-cephfs ([pr#30982](#), Kefu Chai)
- build/ops: rpm: always build ceph-test package ([pr#30188](#), Nathan Cutler)
- ceph-volume: PVolumes.filter shouldn't purge itself ([pr#30806](#), Rishabh Dave)
- ceph-volume: VolumeGroups.filter shouldn't purge itself ([pr#30808](#), Rishabh Dave)
- ceph-volume: add Ceph's device id to inventory ([pr#31211](#), Sebastian Wagner)
- ceph-volume: api/lvm: check if list of LVs is empty ([pr#31229](#), Rishabh Dave)
- ceph-volume: assume msgrV1 for all branches containing mimic ([pr#31615](#), Jan Fajerski)
- ceph-volume: batch functional idempotency test fails since message is now on stderr ([pr#29688](#), Jan Fajerski)
- ceph-volume: broken assertion errors after pytest changes ([pr#28948](#), Alfredo Deza)
- ceph-volume: do not fail when trying to remove crypt mapper ([pr#30555](#), Guillaume Abrioux)
- ceph-volume: does not recognize wal/db partitions created by ceph-disk ([pr#29463](#), Jan Fajerski)
- ceph-volume: ensure device lists are disjoint ([pr#30334](#), Jan Fajerski)
- ceph-volume: extend batch ([issue#40919](#), [pr#29243](#), Andrew Schoen, Jan Fajerski, Sébastien Han, Volker Theile)
- ceph-volume: fix stderr failure to decode/encode when redirected ([pr#30301](#), Alfredo Deza)
- ceph-volume: fix warnings raised by pytest ([pr#30678](#), Rishabh Dave)

- ceph-volume: implement `__format__` in `Size` to format sizes in py3 ([pr#30333](#), Jan Fajerski)
- ceph-volume: look for rotational data in `lsblk` ([pr#26991](#), Andrew Schoen)
- ceph-volume: `lvm.activate`: Return an error if WAL/DB devices absent ([pr#29039](#), David Casier)
- ceph-volume: `lvm.zap` fix cleanup for db partitions ([issue#40664](#), [pr#30303](#), Dominik Csapak)
- ceph-volume: minor optimizations related to class `Volumes`'s use ([pr#30096](#), Rishabh Dave)
- ceph-volume: miscellaneous backports ([pr#31227](#), Mohamad Gebai, Andrew Schoen)
- ceph-volume: missing string substitution when reporting mounts ([issue#40977](#), [pr#29350](#), Shyukri Shyukriev)
- ceph-volume: more mimic backports ([pr#29631](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: more missing mimic backports ([pr#31362](#), Mohamad Gebai, Kefu Chai)
- ceph-volume: pre-install python-apt and its variants before test runs ([pr#30295](#), Alfredo Deza)
- ceph-volume: prints errors to `stdout` with `-format json` ([issue#38548](#), [pr#29507](#), Jan Fajerski)
- ceph-volume: prints log messages to `stdout` ([pr#29602](#), Jan Fajerski, Alfredo Deza, Kefu Chai)
- ceph-volume: replace `testinfra` command with `py.test` ([pr#28930](#), Alfredo Deza)
- ceph-volume: simple functional tests drop test for `lvm zap` ([pr#29661](#), Jan Fajerski)
- ceph-volume: simple: when 'type' file is not present activate fails ([pr#29417](#), Jan Fajerski, Alfredo Deza)
- ceph-volume: tests add a sleep in `tox` for slow OSDs after booting ([pr#28947](#), Alfredo Deza)
- ceph-volume: tests set the `noninteractive` flag for Debian ([pr#29900](#), Alfredo Deza)
- ceph-volume: update testing playbook 'deploy.yml' ([pr#29074](#), Andrew Schoen, Guillaume Abrioux)
- ceph-volume: use the OSD identifier when reporting success ([pr#29770](#), Alfredo Deza)

- ceph-volume: zap always skips block.db, leaves them around ([issue#40664](#), [pr#30306](#), Alfredo Deza)
- ceph\_detect\_init: Add support for ALT Linux ([pr#27028](#), Andrey Bychkov)
- cephfs: MDSTableServer.cc: 83: FAILED assert(version == tid) ([issue#39212](#), [issue#38835](#), [pr#29222](#), "Yan, Zheng")
- cephfs: avoid map been inserted by mistake ([pr#29833](#), XiaoGuoDong2019)
- cephfs: ceph-fuse: client hang because its bad session PipeConnection to mds ([issue#39305](#), [issue#39685](#), [pr#29200](#), Guan yunfei)
- cephfs: client: EINVAL may be returned when offset is 0 ([pr#30932](#), wenpengLi)
- cephfs: client: \_readdir\_cache\_cb() may use the readdir\_cache already clear ([issue#41148](#), [pr#30933](#), huanwen ren)
- cephfs: client: add procession of SEEK\_HOLE and SEEK\_DATA in lseek ([pr#30918](#), Shen Hang)
- cephfs: client: bump ll\_ref from int32 to uint64\_t ([pr#29187](#), Xiaoxi CHEN)
- cephfs: client: ceph.dir.rctime xattr value incorrectly prefixes 09 to the nanoseconds component ([issue#40168](#), [pr#28501](#), David Disseldorp)
- cephfs: client: fix bad error handling in \_lookup\_parent ([issue#40085](#), [pr#29609](#), Jeff Layton)
- cephfs: client: nfs-ganesha with cephfs client, removing dir reports not empty ([issue#40746](#), [pr#30443](#), Peng Xie)
- cephfs: client: return -EIO when sync file which unsafe reqs have been dropped ([issue#40877](#), [pr#30241](#), simon gao)
- cephfs: client: set snapdir's link count to 1 ([pr#30108](#), "Yan, Zheng")
- cephfs: client: support the fallocate() when fuse version >= 2.9 ([issue#40615](#), [pr#30228](#), huanwen ren)
- cephfs: client: unlink dentry for inode with llref=0 ([issue#40960](#), [pr#29479](#), Xiaoxi CHEN)
- cephfs: fix a memory leak ([pr#29915](#), XiaoGuoDong2019)
- cephfs: setattr on snap inode stuck ([issue#40437](#), [pr#29230](#), "Yan, Zheng")
- cephfs: kcephfs TestClientLimits.test\_client\_pin fails with client caps fell below min ([issue#38270](#), [issue#38687](#), [pr#29211](#), "Yan, Zheng")
- cephfs: mds: Fix duplicate client entries in eviction list ([pr#30950](#), Sidharth Anupkrishnan)

- cephfs: mds: avoid sending too many osd requests at once after mds restarts ([issue#40042](#), [issue#40028](#), [pr#28650](#), simon gao)
- cephfs: mds: behind on trimming and [dentry] was purgeable but no longer is! ([issue#39223](#), [issue#38679](#), [pr#29224](#), "Yan, Zheng")
- cephfs: mds: cannot switch mds state from standby-replay to active ([issue#40213](#), [pr#29232](#), "Yan, Zheng", simon gao)
- cephfs: mds: change how mds revoke stale caps ([issue#38043](#), [issue#17854](#), [pr#28585](#), "Yan, Zheng", Rishabh Dave)
- cephfs: mds: check dir fragment to split dir if mkdir makes it oversized ([issue#39689](#), [pr#28381](#), Erqi Chen)
- cephfs: mds: cleanup unneeded client\_snap\_caps when splitting snap inode ([issue#39987](#), [pr#30234](#), "Yan, Zheng")
- cephfs: mds: delay exporting directory whose pin value exceeds max rank id ([issue#40603](#), [pr#29940](#), Zhi Zhang)
- cephfs: mds: destroy reconnect msg when it is from non-existent session to avoid memory leak ([issue#40588](#), [pr#28796](#), Shen Hang)
- cephfs: mds: evict an unresponsive client only when another client wants its caps ([pr#30239](#), Rishabh Dave)
- cephfs: mds: fix SnapRealm::resolve\_snapname for long name ([issue#39472](#), [pr#28186](#), "Yan, Zheng")
- cephfs: mds: fix corner case of replaying open sessions ([pr#28579](#), "Yan, Zheng")
- cephfs: mds: high debug logging with many subtrees is slow ([issue#38875](#), [pr#29219](#), Rishabh Dave)
- cephfs: mds: make MDSIOContextBase delete itself when shutting down ([pr#30417](#), Xuehan Xu)
- cephfs: mds: mds\_cap\_revoke\_eviction\_timeout is not used to initialize Server::cap\_revoke\_eviction\_timeout ([issue#38844](#), [issue#39210](#), [pr#29220](#), simon gao)
- cephfs: mds: output lock state in format dump ([issue#39669](#), [issue#39645](#), [pr#28274](#), Zhi Zhang)
- cephfs: mds: remove cache drop admin socket command ([issue#38020](#), [issue#38099](#), [pr#29210](#), Patrick Donnelly)
- cephfs: mds: reset heartbeat during long-running loops in recovery ([issue#40222](#), [pr#28918](#), "Yan, Zheng")

- cephfs: mds: stopping MDS with a large cache (40+GB) causes it to miss heartbeats ([issue#38022](#), [issue#38129](#), [issue#37723](#), [issue#38131](#), [pr#28452](#), Patrick Donnelly)
- cephfs: mds: there is an assertion when calling Beacon::shutdown() ([issue#39215](#), [issue#38822](#), [pr#29223](#), huanwen ren)
- cephfs: mount.ceph.c: do not pass nofail to the kernel ([issue#39233](#), [pr#28090](#), Kenneth Waegeman)
- cephfs: mount.ceph: properly handle -o strictatime ([pr#30240](#), Jeff Layton)
- cephfs: mount: key parsing fail when doing a remount ([issue#40165](#), [pr#29225](#), Luis Henriques)
- cephfs: pybind: added lseek() ([issue#39679](#), [pr#28337](#), Xiaowei Chu)
- cephfs: test\_volume\_client: fix test\_put\_object\_versioned() ([issue#39405](#), [issue#39510](#), [pr#30236](#), Rishabh Dave)
- common/ceph\_context: avoid unnecessary wait during service thread shutdown ([pr#31096](#), Jason Dillaman)
- common/options.cc: Lower the default value of osd\_deep\_scrub\_large\_omap\_object\_key\_threshold ([pr#29174](#), Neha Ojha)
- common/util: handle long lines in /proc/cpuinfo ([issue#39475](#), [issue#38296](#), [pr#28206](#), Sage Weil)
- common: Keyrings created by ceph auth get are not suitable for ceph auth import ([issue#22227](#), [issue#40547](#), [pr#28741](#), Kefu Chai)
- common: data race in OutputDataSocket ([issue#40268](#), [issue#40188](#), [pr#29201](#), Casey Bodley)
- common: parse ISO 8601 datetime format ([issue#40088](#), [pr#28326](#), Sage Weil)
- core: .mgrstat failed to decode mgrstat state; luminous dev version? ([issue#38852](#), [issue#38839](#), [pr#29249](#), Sage Weil)
- core: Better default value for osd\_snap\_trim\_sleep ([pr#29732](#), Neha Ojha)
- core: Health warnings on long network ping times ([issue#40640](#), [issue#40586](#), [pr#30225](#), xie xingguo, David Zafman)
- core: ceph daemon mon.a config set mon\_health\_to\_clog false cause leader mon assert ([issue#39625](#), [pr#29741](#), huangjun)
- core: crc cache should be invalidated when posting preallocated rx buffers ([issue#38437](#), [pr#29247](#), Ilya Dryomov)
- core: lazy omap stat collection ([pr#29189](#), Brad Hubbard)

- core: mon, osd: parallel clean\_pg\_upmaps ([issue#40104](#), [issue#40230](#), [pr#28619](#), xie xingguo)
- core: mon,osd: limit MOSDMap messages by size as well as map count ([issue#38277](#), [issue#38040](#), [pr#29242](#), Sage Weil)
- core: mon/AuthMonitor: fix initial creation of rotating keys ([issue#40634](#), [pr#30181](#), Sage Weil)
- core: mon/MDSMonitor: use stringstream instead of dout for mds repaired ([issue#40472](#), [pr#30235](#), Zhi Zhang)
- core: mon/MgrMonitor: fix null deref when invalid formatter is specified ([pr#29593](#), Sage Weil)
- core: mon/OSDMonitor.cc: better error message about min\_size ([pr#29618](#), Neha Ojha)
- core: mon/OSDMonitor: trim not-longer-exist failure reporters ([pr#30903](#), NancySu05)
- core: mon: C\_AckMarkedDown has not handled the Callback Arguments ([pr#30213](#), NancySu05)
- core: mon: ensure prepare\_failure() marks no\_reply on op ([pr#30481](#), Joao Eduardo Luis)
- core: mon: paxos: introduce new reset\_pending\_committing\_finishers for safety ([issue#39744](#), [issue#39484](#), [pr#28540](#), Greg Farnum)
- core: mon: show pool id in pool ls command ([issue#40287](#), [pr#30485](#), Chang Liu)
- core: osd beacon sometimes has empty pg list ([issue#40464](#), [issue#40377](#), [pr#29253](#), Sage Weil)
- core: osd/OSD.cc: make osd bench description consistent with parameters ([issue#39374](#), [issue#39006](#), [pr#28097](#), Neha Ojha)
- core: osd/OSDCap: Check for empty namespace ([issue#40835](#), [pr#30214](#), Brad Hubbard)
- core: osd/OSDMap: Replace get\_out\_osds with get\_out\_existing\_osds ([issue#39422](#), [issue#39154](#), [pr#28142](#), Brad Hubbard)
- core: osd/OSDMap: do not trust partially simplified pg\_upmap\_item ([pr#30898](#), xie xingguo)
- core: osd/PG: Add PG to large omap log message ([pr#30924](#), Brad Hubbard)
- core: osd/PG: fix last\_complete re-calculation on splitting ([issue#39538](#), [issue#26958](#), [pr#28259](#), xie xingguo)
- core: osd/PeeringState: do not complain about past\_intervals constrained by

- oldest epoch ([pr#30222](#), Sage Weil)
- core: osd/PeeringState: recover\_got - add special handler for empty log ([pr#30895](#), xie xingguo)
  - core: osd/PrimaryLogPG: Avoid accessing destroyed references in finish\_degr... ([pr#30291](#), Tao Ning)
  - core: osd/PrimaryLogPG: skip obcs that don't exist during backfill scan\_range ([pr#31029](#), Sage Weil)
  - core: osd/PrimaryLogPG: update oi.size on write op implicitly truncating ob... ([pr#30275](#), xie xingguo)
  - core: osd: Better error message when OSD count is less than osd\_pool\_default\_size ([issue#38617](#), [pr#30180](#), Kefu Chai, Sage Weil, zjh)
  - core: osd: Don't evict after a flush if intersecting scrub range ([issue#38840](#), [issue#39518](#), [pr#28232](#), David Zafman)
  - core: osd: Don't include user changeable flag in snaptrim related assert ([issue#38124](#), [issue#39698](#), [pr#28202](#), David Zafman)
  - core: osd: Fix for compatibility of encode/decode of osd\_stat\_t ([pr#31275](#), Kefu Chai, David Zafman)
  - core: osd: Include dups in copy\_after() and copy\_up\_to() ([issue#39304](#), [pr#28089](#), David Zafman)
  - core: osd: Output Base64 encoding of CRC header if binary data present ([issue#39737](#), [pr#28503](#), David Zafman)
  - core: osd: Remove unused osdmap flags full, nearfull from output ([pr#30901](#), David Zafman)
  - core: osd: clear PG\_STATE\_CLEAN when repair object ([pr#30243](#), Zengran Zhang)
  - core: osd: fix build\_incremental\_map\_msg ([issue#38282](#), [pr#31236](#), Sage Weil)
  - core: osd: make project\_pg\_history handle concurrent osdmap publish ([issue#26970](#), [pr#29976](#), Sage Weil)
  - core: osd: merge replica log on primary need according to replica log's crt ([pr#30916](#), Zengran Zhang)
  - core: osd: pg stuck in backfill\_wait with plenty of disk space ([issue#38034](#), [pr#28201](#), xie xingguo, David Zafman)
  - core: osd: report omap/data/metadata usage ([issue#40639](#), [pr#28852](#), Sage Weil)
  - core: osd: rollforward may need to mark pglog dirty ([issue#40403](#), [pr#31035](#), Zengran Zhang)

- core: osd: scrub error on big objects; make bluestore refuse to start on big objects ([pr#30784](#), David Zafman, Sage Weil)
- core: osd: take heartbeat\_lock when calling heartbeat() ([issue#39513](#), [issue#39439](#), [pr#28220](#), Sage Weil)
- core: osds allows to partially start more than N+2 ([issue#38206](#), [issue#38076](#), [pr#29241](#), Sage Weil)
- core: should report EINVAL in ErasureCode::parse() if m<=0 ([issue#38682](#), [issue#38751](#), [pr#28995](#), Sage Weil)
- core: should set EPOLLET flag on del\_event() ([issue#38856](#), [pr#29250](#), Roman Penyaev)
- doc/ceph-fuse: mention -k option in ceph-fuse man page ([pr#30936](#), Rishabh Dave)
- doc/rbd: s/guess/xml/ for codeblock lexer ([pr#31090](#), Kefu Chai)
- doc/rgw: document use of 'realm pull' instead of 'period pull' ([issue#39655](#), [pr#30131](#), Casey Bodley)
- doc: Document behaviour of fsync-after-close ([issue#24641](#), [pr#29765](#), Jos Collin, Jeff Layton)
- doc: Object Gateway multisite document read-only argument error ([issue#40497](#), [pr#29289](#), Chenjiong Deng)
- doc: default values for mon\_health\_to\_clog\_\* were flipped ([pr#30227](#), James McClune)
- doc: describe metadata\_heap cleanup ([issue#18174](#), [pr#30070](#), Dan van der Ster)
- doc: fix rgw\_ldap\_dnattr username token ([pr#30099](#), Thomas Kriechbaumer)
- doc: rgw: CreateBucketConfiguration for s3 PUT Bucket request ([issue#39602](#), [issue#39597](#), [pr#29257](#), Casey Bodley)
- doc: update bluestore cache settings and clarify data fraction ([issue#39522](#), [pr#31258](#), Jan Fajerski)
- doc: wrong value of usage log default in logging section ([issue#37891](#), [issue#37856](#), [pr#29014](#), Abhishek Lekshmanan)
- filestore: assure sufficient leaves in pre-split ([issue#39390](#), [pr#30182](#), Jeegn Chen)
- krbd: avoid udev netlink socket overrun and retry on transient errors from udev\_enumerate\_scan\_devices() ([pr#31322](#), Ilya Dryomov, Adam C. Emerson)
- krbd: fix rbd map hang due to udev return subsystem unordered ([issue#39089](#), [pr#30176](#), Zhi Zhang)

- mgr/balancer: fix fudge ([pr#28399](#), xie xingguo)
- mgr/balancer: python3 compatibility issue ([pr#31013](#), Mykola Golub)
- mgr/balancer: restrict automatic balancing to specific weekdays ([pr#26499](#), xie xingguo)
- mgr/crash: fix python3 invalid syntax problems ([pr#29029](#), Ricardo Dias)
- mgr/dashboard: Fix run-frontend-e2e-tests.sh ([issue#40707](#), [pr#28954](#), Kiefer Chang, Tiago Melo)
- mgr/dashboard: Fix various RGW issues ([pr#28210](#), Volker Theile)
- mgr/dashboard: RGW proxy can't handle self-signed SSL certificates ([pr#30543](#), Volker Theile)
- mgr/dashboard: cephfs multimds graphs stack together ([issue#40660](#), [pr#28911](#), Kiefer Chang)
- mgr/localpool: pg\_num is an int arg to 'osd pool create' ([pr#30447](#), Sage Weil)
- mgr/prometheus: Cast collect\_timeout (scrape\_interval) to float ([pr#31108](#), Ben Meekhof)
- mgr/prometheus: replace whitespaces in metrics' names ([issue#39458](#), [pr#28165](#), Alfonso Martínez)
- mgr/telemetry: Ignore crashes in report when module not enabled ([pr#30846](#), Wido den Hollander)
- mgr: DaemonServer::handle\_conf\_change - broken locking ([issue#38899](#), [issue#38963](#), [pr#29197](#), xie xingguo)
- mgr: deadlock ([issue#39040](#), [issue#39426](#), [pr#28161](#), xie xingguo)
- mgr: do not reset reported if a new metric is not collected ([pr#30391](#), Ilsoo Byun)
- radosgw-admin: bucket sync status not 'caught up' during full sync ([issue#40806](#), [pr#30170](#), Casey Bodley)
- rbd-mirror: cannot restore deferred deletion mirrored images ([pr#30828](#), Jason Dillaman, Mykola Golub)
- rbd-mirror: clear out bufferlist prior to listing mirror images ([issue#39461](#), [issue#39407](#), [pr#28123](#), Jason Dillaman)
- rbd-mirror: don't overwrite status error returned by replay ([pr#29872](#), Mykola Golub)
- rbd-mirror: handle duplicates in image sync throttler queue ([issue#40519](#),

[issue#40593](#), [pr#28815](#), Mykola Golub)

- rbd-mirror: ignore errors relating to parsing the cluster config file ([pr#30117](#), Jason Dillaman)
- rbd/action: fix error getting positional argument ([issue#40095](#), [pr#29294](#), songweibin)
- rbd/tests: avoid hexdump skip and length options in krbd test ([pr#30569](#), Ilya Dryomov)
- rbd: Reduce log level for cls/journal and cls/rbd expected errors ([issue#40865](#), [pr#29565](#), Jason Dillaman)
- rbd: filter out group/trash snapshots from snap\_list ([issue#38538](#), [issue#39186](#), [pr#28138](#), songweibin, Jason Dillaman)
- rbd: journal: properly advance read offset after skipping invalid range ([pr#28814](#), Mykola Golub)
- rbd: librbd: add missing shutdown states to managed lock helper ([issue#38387](#), [issue#38509](#), [pr#28151](#), Jason Dillaman)
- rbd: librbd: async open/close should free ImageCtx before issuing callback ([issue#39429](#), [issue#39031](#), [pr#28125](#), Jason Dillaman)
- rbd: librbd: avoid dereferencing an empty container during deep-copy ([issue#40368](#), [pr#30177](#), Jason Dillaman)
- rbd: librbd: disable image mirroring when moving to trash ([pr#28150](#), Mykola Golub)
- rbd: librbd: ensure compare-and-write doesn't skip compare after copyup ([issue#38383](#), [issue#38441](#), [pr#28133](#), Ilya Dryomov)
- rbd: librbd: properly handle potential object map failures ([issue#39952](#), [issue#36074](#), [pr#30796](#), Jason Dillaman, Mykola Golub)
- rbd: librbd: properly track in-flight flush requests ([issue#40573](#), [pr#28770](#), Jason Dillaman)
- rbd: librbd: race condition possible when validating RBD pool ([issue#38500](#), [issue#38563](#), [pr#28139](#), Jason Dillaman)
- rbd: use the ordered throttle for the export action ([issue#40435](#), [pr#30178](#), Jason Dillaman)
- restful: Query nodes\_by\_id for items ([pr#31273](#), Boris Ranto)
- rgw admin: disable stale instance delete in a multiste env ([pr#30340](#), Abhishek Lekshmanan)

- rgw/OutputDataSocket: append\_output(buffer::list&) says it will (but does not) discard output at data\_max\_backlog ([issue#40178](#), [issue#40351](#), [pr#29279](#), Matt Benjamin)
- rgw/cls: keep issuing bilog trim ops after reset ([issue#40187](#), [pr#30074](#), Casey Bodley)
- rgw/multisite: Don't allow certain radosgw-admin commands to run on non-master zone ([issue#39548](#), [pr#30133](#), Shilpa Jagannath)
- rgw/rgw\_op: Remove get\_val from hotpath via legacy options ([pr#30141](#), Mark Nelson)
- rgw: Add support for -bypass-gc flag of radosgw-admin bucket rm command in RGW Multi-site ([issue#39748](#), [issue#24991](#), [pr#29262](#), Casey Bodley)
- rgw: Don't crash on copy when metadata directive not supplied ([issue#40416](#), [pr#29500](#), Adam C. Emerson)
- rgw: Fix bucket versioning vs. swift metadata bug ([pr#30140](#), Marcus Watts)
- rgw: Fix rgw decompression log-print ([pr#30156](#), Han Fengzhe)
- rgw: Multisite sync corruption for large multipart obj ([issue#40144](#), [pr#29273](#), Casey Bodley, Tianshan Qu, Xiaoxi CHEN)
- rgw: RGWCoroutine::call(nullptr) sets retcode=0 ([pr#30159](#), Casey Bodley)
- rgw: Return tenant field in bucket\_stats function ([issue#40038](#), [pr#28209](#), Volker Theile)
- rgw: S3 policy evaluated incorrectly ([issue#38638](#), [issue#39274](#), [pr#29255](#), Pritha Srivastava)
- rgw: Save an unnecessary copy of RGWEnv ([pr#29483](#), Mark Kogan)
- rgw: Swift interface: server side copy fails if object name contains '?' ([issue#27217](#), [issue#40128](#), [pr#29267](#), Casey Bodley)
- rgw: TempURL should not allow PUTs with the X-Object-Manifest ([issue#40133](#), [issue#20797](#), [pr#28711](#), Radoslaw Zarzynski)
- rgw: abort multipart fix ([pr#29016](#), J. Eric Ivancich)
- rgw: asio: check the remote endpoint before processing requests ([pr#30977](#), Abhishek Lekshmanan)
- rgw: conditionally allow builtin users with non-unique email addresses ([issue#40089](#), [issue#40507](#), [pr#28716](#), Matt Benjamin)
- rgw: data/bilogs are trimmed when no peers are reading them ([issue#39487](#), [pr#30130](#), Casey Bodley)

- rgw: datalog/mdlog trim commands loop until done ([pr#30868](#), Casey Bodley)
- rgw: do necessary checking of website configuration ([issue#40678](#), [pr#30980](#), Enming Zhang)
- rgw: don't throw when accept errors are happening on frontend ([pr#30154](#), Yuval Lifshitz)
- rgw: fix CreateBucket with BucketLocation parameter failed under default zonegroup ([pr#30171](#), Enming Zhang)
- rgw: fix bucket may redundantly list keys after BI\_PREFIX\_CHAR ([issue#40147](#), [issue#39984](#), [pr#28409](#), Casey Bodley, Tianshan Qu)
- rgw: fix cls\_bucket\_list\_unordered() partial results ([pr#30253](#), Mark Kogan)
- rgw: fix data sync start delay if remote haven't init data\_log ([pr#30510](#), Tianshan Qu)
- rgw: fix drain handles error when deleting bucket with bypass-gc option ([pr#29984](#), dongdong tao)
- rgw: fix list bucket with delimiter wrongly skip some special keys ([issue#40905](#), [pr#30168](#), Tianshan Qu)
- rgw: fix list versions starts with version\_id=null ([pr#30775](#), Tianshan Qu)
- rgw: fix potential realm watch lost ([issue#40991](#), [pr#30167](#), Tianshan Qu)
- rgw: fix race b/w bucket reshards and ops waiting on reshards completion ([pr#29139](#), J. Eric Ivancich)
- rgw: fix refcount tags to match and update object's idtag ([pr#30891](#), J. Eric Ivancich)
- rgw: fixed "unrecognized arg" error when using "radosgw-admin zone rm" ([pr#30172](#), Hongang Chen)
- rgw: gc remove tag after all sub io finish ([issue#40903](#), [pr#30173](#), Tianshan Qu)
- rgw: housekeeping of reset stats operation in radosgw-admin and cls back-end ([pr#30165](#), J. Eric Ivancich)
- rgw: increase beast parse buffer size to 64k ([pr#30450](#), Casey Bodley)
- rgw: ldap auth: S3 auth failure should return InvalidAccessKeyId ([pr#30652](#), Matt Benjamin)
- rgw: make dns hostnames matching case insensitive ([issue#40995](#), [pr#30166](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: mitigate bucket list with max-entries excessively high ([pr#30134](#), J. Eric

Ivancich)

- rgw: multisite: 'radosgw-admin bucket sync status' should call `syncs_from(source.name)` instead of `id` ([issue#40022](#), [issue#40141](#), [pr#29270](#), Casey Bodley)
- rgw: multisite: RGWListBucketIndexesCR for data full sync needs pagination ([issue#39551](#), [issue#40354](#), [pr#29284](#), Shilpa Jagannath)
- rgw: multisite: data sync loops back to the start of the datalog after reaching the end ([issue#39033](#), [issue#39074](#), [pr#29021](#), Casey Bodley)
- rgw: multisite: mismatch of bucket creation times from List Buckets ([issue#39635](#), [issue#39734](#), [pr#28483](#), Casey Bodley)
- rgw: multisite: overwrites in versioning-suspended buckets fail to sync ([issue#38080](#), [issue#37792](#), [pr#29017](#), Casey Bodley)
- rgw: multisite: period pusher gets 403 Forbidden against other zonegroups ([issue#39415](#), [issue#39287](#), [pr#29256](#), Casey Bodley)
- rgw: non-existent mdlog failures logged at level 0 ([issue#38747](#), [issue#40033](#), [pr#28757](#), Abhishek Lekshmanan)
- rgw: perfcounters: add gc retire counter ([pr#30073](#), Matt Benjamin)
- rgw: permit rgw-admin to populate user info by access-key ([pr#30105](#), Matt Benjamin, Marc Koderer)
- rgw: provide admin-friendly reshards status output ([issue#37615](#), [issue#40357](#), [pr#29285](#), Mark Kogan)
- rgw: remove\_olh\_pending\_entries() does not limit the number of xattrs to remove ([issue#39179](#), [issue#39118](#), [pr#28348](#), Casey Bodley)
- rgw: resharding of a versioned bucket causes a bucket stats discrepancy ([issue#39532](#), [pr#28249](#), J. Eric Ivancich)
- rgw: return ERR\_NO\_SUCH\_BUCKET early while evaluating bucket policy ([issue#38420](#), [issue#39697](#), [pr#28422](#), Abhishek Lekshmanan)
- rgw: rgw\_file: all directories are virtual with respect to contents ([issue#40262](#), [issue#40204](#), [pr#28887](#), Matt Benjamin)
- rgw: set null version object issues ([issue#36763](#), [issue#40360](#), [pr#29288](#), Tianshan Qu)
- rgw: support delimiter longer than one symbol ([issue#39989](#), [issue#38776](#), [pr#29018](#), Tianshan Qu, Matt Benjamin)
- rgw: swift object expiry fails when a bucket reshards ([issue#39741](#), [pr#29258](#),

Casey Bodley, Abhishek Lekshmanan, J. Eric Ivancich)

- rgw: swift: refrain from corrupting static large objects when using nginx as a GET cache ([pr#30135](#), Andrey Groshev)
- rgw: the Multi-Object Delete operation of S3 API wrongly handles the Code response element ([issue#18241](#), [issue#40136](#), [pr#29268](#), Radoslaw Zarzynski)
- rgw: update resharding documentation ([issue#39047](#), [pr#29020](#), J. Eric Ivancich)
- rgw\_file: fix invalidation of top-level directories ([issue#40215](#), [pr#29276](#), Matt Benjamin)
- rgw\_file: advance\_mtime() should consider namespace expiration ([issue#40415](#), [pr#30660](#), Matt Benjamin)
- rgw\_file: fix readdir eof() calc-caller stop implies !eof and introduce fast S3 Unix stats (immutable) ([issue#40375](#), [issue#40456](#), [pr#30077](#), Matt Benjamin)
- rgw\_file: include tenant when hashing bucket names ([issue#40225](#), [issue#40118](#), [pr#29277](#), Matt Benjamin)
- rgw\_file: readdir: do not construct markers w/leading '/' ([pr#30157](#), Matt Benjamin)
- rgw\_file: save etag and acl info in setattr ([issue#39229](#), [pr#28073](#), Tao Chen)
- rpm: missing dependency on python34-ceph-argparse from python34-cephfs (and others?) ([issue#24918](#), [issue#24919](#), [issue#37613](#), [pr#27949](#), Kefu Chai)
- tests: cls\_rbd: removed mirror peer pool test cases ([pr#31485](#), Jason Dillaman)
- tests: librbd: set nbd timeout due to newer kernels defaulting it on ([pr#30424](#), Jason Dillaman)
- tests: ceph-disk: use a Python2.7 compatible version of pytest ([pr#31254](#), Alfredo Deza)
- tests: rgw: don't use ceph-ansible in s3a-hadoop suite ([issue#39706](#), [pr#30069](#), Casey Bodley)
- tests/workunits/rbd: wait for rbd-nbd unmap to complete ([issue#39598](#), [issue#39674](#), [pr#28310](#), Jason Dillaman)
- tests: fix issues in vstart runner ([pr#28208](#), Volker Theile)
- tests: limit loops waiting for force-backfill/force-recovery to happen ([issue#38351](#), [issue#38309](#), [pr#29245](#), David Zafman)
- tests: remove s3tests ! ([pr#31640](#), Yuri Weinstein)
- tests: cephfs: TestMisc.test\_evict\_client fails ([issue#40219](#), [pr#29228](#), "Yan,

Zheng")

- tests: do not take ceph.conf.template from ceph/teuthology.git ([pr#30841](#), Sage Weil)
- tests: ignore expected MDS\_CLIENT\_LATE\_RELEASE warning ([issue#40968](#), [pr#29812](#), Patrick Donnelly)
- tests: install python3-cephfs for fs suite ([pr#31285](#), Kefu Chai)
- tests: kclient unmount hangs after file system goes down ([issue#38709](#), [issue#38677](#), [pr#29218](#), Patrick Donnelly)
- tests: krbd\_msgr\_segments.t: filter lvcreate output ([pr#31324](#), Ilya Dryomov)
- tests: make get\_mon\_status use mon addr ([pr#31461](#), Sage Weil, Nathan Cutler)
- tests: make: \*\*\* [hello\_world\_cpp] Error 127 in rados ([issue#40320](#), [pr#29203](#), Kefu Chai)
- tests: qa/standalone/scrub/osd-scrub-snaps.sh sometimes fails ([issue#40179](#), [issue#40078](#), [pr#29251](#), David Zafman)
- tests: qa/tasks/ceph.py: pass cluster\_name to get\_mons ([pr#31424](#), Nathan Cutler)
- tests: qa/workunits/rbd: stress test "rbd mirror pool status -verbose" ([pr#29873](#), Mykola Golub)
- tests: remove "1node" and "systemd" tests as ceph-deploy is not actively developed ([pr#28457](#), Yuri Weinstein)
- tests: sleep briefly after resetting kclient ([pr#29751](#), Patrick Donnelly)
- tests: test\_volume\_client: print python version correctly ([issue#40317](#), [issue#40184](#), [pr#29208](#), Lianne)
- tests: use curl in wait\_for\_radosgw() in util/rgw.py ([pr#28668](#), Ali Maredia)
- tests: use hard\_reset to reboot kclient ([issue#37681](#), [pr#30233](#), Patrick Donnelly)
- tests: whitelisted 'application not enabled' ([pr#28389](#), Yuri Weinstein)
- tools/rados: list objects in a pg ([issue#36732](#), [pr#30893](#), Vikhyat Umrao, Li Wang)
- tools/rbd-ggate: close log before running postfork ([pr#30121](#), Willem Jan Withagen)
- tools: Add clear-data-digest command to objectstore tool ([issue#37749](#), [pr#29196](#), Li Yichao)
- tools: ceph-objectstore-tool can't remove head with bad snapset ([pr#30081](#), David Zafman)

- tools: ceph-objectstore-tool: return 0 if incmap is sane ([pr#31659](#), Kefu Chai)
- tools: ceph-objectstore-tool: update-mon-db: do not fail if incmap is missing ([pr#30979](#), Kefu Chai)
- tools: crushtool crash on Fedora 28 and newer ([issue#39174](#), [issue#39311](#), [pr#27986](#), Brad Hubbard)

## v13.2.6 Mimic

---

This is the sixth bugfix release of the Mimic v13.2.x long term stable release series. We recommend all Mimic users upgrade.

## Notable Changes

---

- Ceph v13.2.6 now packages python bindings for python3.6 instead of python3.4, because EPEL7 recently switched from python3.4 to python3.6 as the native python3. See the announcement <<https://lists.fedoraproject.org/archives/list/epel-announce@lists.fedoraproject.org/message/EGUMKAIMPK2UD5VSHXM53BH2MBDGDWM0/>>\\_ for more details on the background of this change.

## Changelog

---

- cephfs: MDSMonitor: do not assign standby-replay when degraded ([issue#36384](#), [pr#26643](#), Patrick Donnelly)
- ceph-volume: add -all flag to simple activate ([pr#26655](#), Jan Fajerski)
- ceph-volume: use our own testinfra suite for functional testing ([pr#26702](#), Andrew Schoen)
- cli: ability to change file ownership ([issue#38370](#), [pr#26760](#), Sébastien Han)
- cli: better output of 'ceph health detail' ([issue#39266](#), [pr#27847](#), Shen Hang)
- cls/rgw: raise debug level of bi\_log\_iterate\_entries output ([pr#27973](#), Casey Bodley)
- common: ceph\_timer: stop timer's thread when it is suspended ([issue#37766](#), [pr#26583](#), Peng Wang)
- common/str\_map: fix trim() on empty string ([issue#38329](#), [pr#26810](#), Sage Weil)
- core: ENOENT in collection\_move\_rename on EC backfill target ([issue#36739](#), [pr#27943](#), Neha Ojha)
- core: Fix recovery and backfill priority handling ([issue#38041](#), [pr#27081](#), David

Zafman)

- crush: add root\_bucket to identify underfull buckets ([issue#38826](#), [pr#27257](#), huangjun)
- crush: backport recent upmap fixes ([issue#37968](#), [issue#38897](#), [issue#37940](#), [pr#27963](#), xie xingguo)
- crush/CrushWrapper: ensure crush\_choose\_arg\_map.size == max\_buckets ([issue#38664](#), [pr#27082](#), Sage Weil)
- doc: Fix incorrect mention of 'osd\_deep\_mon\_scrub\_interval' ([pr#26860](#), Ashish Singh)
- doc: Minor rados related documentation fixes ([issue#38896](#), [pr#27188](#), David Zafman)
- doc: osd\_recovery\_priority is not documented (but osd\_recovery\_op\_priority is) ([issue#23999](#), [pr#26901](#), David Zafman)
- doc/radosgw: Document mappings of S3 Operations to ACL grants ([issue#38523](#), [pr#26968](#), Adam C. Emerson)
- doc/rgw: document placement target configuration ([issue#24508](#), [pr#27032](#), Casey Bodley)
- doc: Update bluestore config docs - fix typo (as -> has) ([pr#27845](#), Yaniv Kaul)
- doc: updated reference link for log based PG ([issue#38465](#), [pr#26829](#), James McClune)
- include/intarith: enforce the same type for p2\*() arguments ([pr#27318](#), Ilya Dryomov)
- librbd: avoid aggregate-initializing any static\_visitor ([issue#38659](#), [pr#27041](#), Willem Jan Withagen)
- librbd: avoid aggregate-initializing IsWriteOpVisitor ([issue#38660](#), [pr#27039](#), Willem Jan Withagen)
- mds: drop reconnect message from non-existent session ([issue#39026](#), [pr#27916](#), Shen Hang)
- mds: inode filtering on 'dump cache' asok ([issue#11172](#), [pr#27058](#), dongdong tao)
- mds/server: check directory split after rename ([issue#38994](#), [pr#27917](#), Shen Hang)
- mds: wait for client to release shared cap when re-acquiring xlock ([issue#38491](#), [pr#27023](#), "Yan, Zheng")
- mgr/balancer: blame if upmap won't actually work ([issue#38780](#), [pr#26497](#), xie xingguo)

- mgr/BaseMgrModule: drop GIL for ceph\_send\_command ([issue#38537](#), [pr#26833](#), Sage Weil)
- mgr: crashdump feature backport ([pr#24639](#), Noah Watkins, Sage Weil, Dan Mick)
- mgr/dashboard: fix for using '::' on hosts without ipv6 ([issue#38575](#), [pr#26750](#), Noah Watkins)
- mgr/dashboard: Manager should complain about wrong dashboard certificate ([issue#24453](#), [pr#27747](#), Volker Theile, Ricardo Dias)
- mgr/dashboard: Search broken for entries with null values ([issue#38583](#), [pr#26944](#), Patrick Nawracay)
- mgr/dashboard: show I/O stats in Pool list ([pr#27053](#), Alfonso Martínez)
- mgr/dashboard: Update npm packages ([issue#39080](#), [pr#26670](#), Tiago Melo)
- mgr/dashboard: Use human readable units on the OSD I/O graphs ([issue#25075](#), [pr#27558](#), Tiago Melo)
- mgr: drop GIL in get\_config ([pr#26612](#), John Spray)
- mgr: enable inter-module calls ([pr#27638](#), John Spray)
- mgr/prometheus: add interface and objectstore to osd metadata ([pr#26537](#), Jan Fajerski, Konstantin Shalygin)
- mgr/PyModule: put mgr\_module\_path first in sys.path ([issue#38469](#), [pr#26777](#), Tim Serong)
- mon/OSDMonitor: fix osd boot check ([pr#27351](#), Sage Weil)
- mon/OSDMonitor: further improve prepare\_command\_pool\_set E2BIG error message ([issue#39353](#), [pr#27647](#), Nathan Cutler)
- msg: output peer address when detecting bad CRCs ([issue#39367](#), [pr#27860](#), Greg Farnum)
- multisite: bucket full sync does not handle delete markers ([issue#38007](#), [pr#26194](#), Casey Bodley)
- multisite: rgw\_data\_sync\_status json decode failure breaks automated datalog trimming ([issue#38373](#), [pr#26615](#), Casey Bodley)
- os/bluestore: backport new bitmap allocator ([pr#26983](#), Igor Fedotov, Sage Weil)
- os/bluestore: bitmap allocator might fail to return contiguous chunk despite having enough space ([pr#27298](#), Igor Fedotov)
- os/bluestore: call fault\_range properly prior to looking for blob to ... ([pr#27570](#), Igor Fedotov)

- os/bluestore: fix improper backport for p2 macros for bmap allocator ([pr#27606](#), Igor Fedotov)
- os/bluestore: fix length overflow ([issue#39245](#), [pr#27366](#), Jianpeng Ma)
- os/bluestore: fix out-of-bound access in bmap allocator ([pr#27738](#), Igor Fedotov)
- os/bluestore\_tool: bluefs-bdev-expand: indicate bypassed for main dev ([pr#27447](#), Igor Fedotov)
- osd: FAILED ceph\_assert(attrs || !pg\_log.get\_missing().is\_missing(soid) || (it\_objects != pg\_log.get\_log().objects.end() && it\_objects->second->op == pg\_log\_entry\_t::LOST\_REVERT)) in PrimaryLogPG::get\_object\_context() ([issue#38931](#), [issue#38784](#), [pr#27940](#), xie xingguo)
- osd: fixup OpTracker destruct assert, waiting\_for\_osdmap take ref with OpRequest ([issue#38377](#), [pr#26862](#), linbing)
- osd/PG: discover missing objects when an OSD peers and PG is degraded ([pr#27745](#), Jonas Jelten)
- osd/PGLog.h: print olog\_can\_rollback\_to before deciding to rollback ([issue#38894](#), [pr#27284](#), Neha Ojha)
- osd/PGLog: preserve original\_crt to check rollbackability ([issue#39023](#), [issue#36739](#), [pr#27629](#), Neha Ojha)
- osd/PrimaryLogPG: handle object !exists in handle\_watch\_timeout ([issue#38432](#), [pr#26709](#), Sage Weil)
- osd: process\_copy\_chunk remove obc ref before pg unlock ([issue#38842](#), [pr#27587](#), Zengran Zhang)
- osd: shutdown recovery\_request\_timer earlier ([issue#38945](#), [pr#27938](#), Zengran Zhang)
- pybind/rados: fixed Python3 string conversion issue on get\_fsid ([issue#38381](#), [pr#27259](#), Jason Dillaman)
- rbd: API list\_images() Segmentation fault ([issue#38468](#), [pr#26707](#), songweibin)
- rbd: krbd: return -ETIMEDOUT in polling ([issue#38792](#), [pr#27588](#), Dongsheng Yang)
- rbd\_mirror: don't report error if image replay canceled ([pr#26140](#), Mykola Golub)
- rgw: Adding tcp\_nodelay option to Beast ([issue#34308](#), [pr#27367](#), Or Friedmann)
- rgw admin: add tenant argument to reshard cancel ([issue#38214](#), [pr#27603](#), Abhishek Lekshmanan)
- rgw-admin: fix data sync report for master zone ([issue#38938](#), [pr#27421](#), cfanz)

- rgw: admin: handle delete\_at attr in object stat output ([pr#27828](#), Abhishek Lekshmanan)
- rgw: allow radosgw-admin to list bucket w -allow-unordered ([pr#28096](#), J. Eric Ivancich)
- rgw: beast: set a default port for endpoints ([issue#39000](#), [pr#27661](#), Abhishek Lekshmanan)
- rgw: bucket limit check misbehaves for > max-entries buckets (usually 1000) ([pr#26945](#), Matt Benjamin)
- rgw: bug in versioning concurrent, list and get have consistency issue ([issue#38060](#), [pr#26664](#), Wang Hao)
- rgw: check for non-existent bucket in RGWGetACLs ([issue#38116](#), [pr#26529](#), Matt Benjamin)
- rgw: cls\_bucket\_list\_unordered lists a single shard ([issue#39393](#), [pr#28086](#), Casey Bodley)
- rgw: data sync drains lease stack on lease failure ([issue#38479](#), [pr#26762](#), Casey Bodley)
- rgw: don't crash on missing /etc/mime.types ([issue#38328](#), [pr#27354](#), Casey Bodley)
- rgw: failed to pass test\_bucket\_create\_naming\_bad\_punctuation in s3test ([issue#23587](#), [issue#26965](#), [pr#27666](#), yuliyang, Abhishek Lekshmanan)
- rgw: fix bug of apply default quota, for this create new a user may core using beast ([issue#38847](#), [pr#27335](#), liaoxin01)
- rgw: fix read not exists null version return wrong ([issue#38811](#), [pr#27304](#), Tianshan Qu)
- rgw: Fix S3 compatibility bug when CORS is not found ([issue#37945](#), [pr#27356](#), Nick Janus)
- rgw: GetBucketCORS API returns Not Found error code when CORS configuration does not exist ([issue#26964](#), [pr#27122](#), yuliyang, ashitakasam)
- rgw: get or set realm zonegroup zone should check user's caps for security ([issue#37352](#), [pr#27948](#), yuliyang, Casey Bodley)
- rgw: ldap: fix LDAPAuthEngine::init() when uri !empty() ([issue#38699](#), [pr#27174](#), Matt Benjamin)
- rgw: multiple es related fixes and improvements ([issue#38028](#), [issue#22877](#), [issue#36233](#), [issue#38030](#), [issue#36092](#), [pr#26517](#), Yehuda Sadeh, Abhishek Lekshmanan, Willem Jan Withagen)

- rgw: nfs: skip empty (non-POSIX) path segments ([issue#38744](#), [pr#27179](#), Matt Benjamin)
- rgw: only update last\_trim marker on ENODATA ([issue#38075](#), [pr#26641](#), Casey Bodley)
- rgw: resolve bugs and clean up garbage collection code ([issue#38454](#), [pr#27796](#), J. Eric Ivancich)
- rgw: rgw\_file: use correct secret key to check auth ([issue#37855](#), [pr#26687](#), MinSheng Lin)
- rgw: sse c fixes ([issue#38700](#), [pr#27297](#), Adam Kupczyk, Casey Bodley, Abhishek Lekshmanan)
- rgw: sync module: avoid printing attrs of objects in log ([issue#37646](#), [pr#27029](#), Abhishek Lekshmanan)
- rgw: use chunked encoding to get partial results out faster ([issue#12713](#), [pr#28014](#), Robin H. Johnson)
- rgw: when exclusive lock fails due existing lock, log add'l info ([issue#38171](#), [pr#26553](#), J. Eric Ivancich)
- rgw: when using nfs-ganesha to upload file, rgw es sync module get failed ([issue#36233](#), [pr#27972](#), Abhishek Lekshmanan)
- run-standalone.sh: Need double-quotes to handle | in core\_pattern on all distributions ([issue#38325](#), [pr#26811](#), David Zafman)
- spdk: update to latest spdk-18.05 branch ([pr#27451](#), Kefu Chai)
- test: run-standalone.sh set local library location so mgr can find li... ([issue#38262](#), [pr#26495](#), David Zafman)
- test/store\_test: fix/workaround for BlobReuseOnOverwriteUT and garbageCollection ([pr#27055](#), Igor Fedotov)
- test: Verify a log trim trims the dup\_index ([pr#26578](#), Brad Hubbard)
- tools: ceph-disk/tests: use random unused port for CEPH\_MON ([issue#39066](#), [pr#27228](#), Kefu Chai)
- tools: ceph-objectstore-tool: rename dump-import to dump-export ([issue#39284](#), [pr#27635](#), David Zafman)

## v13.2.5 Mimic

---

This is the fifth bugfix release of the Mimic v13.2.x long term stable release series. We recommend all Mimic users upgrade.

# Notable Changes

---

- This release fixes the pg log hard limit bug that was introduced in 13.2.2, <https://tracker.ceph.com/issues/36686>. A flag called pglog\_hardlimit has been introduced, which is off by default. Enabling this flag will limit the length of the pg log. In order to enable that, the flag must be set by running ceph osd set pglog\_hardlimit after completely upgrading to 13.2.2. Once the cluster has this flag set, the length of the pg log will be capped by a hard limit. Once set, this flag *must not* be unset anymore. In luminous, this feature was introduced in 12.2.11. Users who are running 12.2.11, and want to continue to use this feature, should upgrade to 13.2.5 or later.
- This release also fixes a CVE on civetweb, CVE-2019-3821 where SSL file descriptors were not closed in civetweb in case the initial negotiation fails.
- There have been fixes to RGW dynamic and manual resharding, which no longer leaves behind stale bucket instances to be removed manually. For finding and cleaning up older instances from a reshard a radosgw-admin command reshard stale-instances list and reshard stale-instances rm should do the necessary cleanup. These commands should *not* be used on a multisite setup as the stale instances may be unlikely to be from a reshard and can have consequences. In the next version the admin CLI will prevent this command to be run on a multisite cluster, however for the current release users are urged not to use the delete command on a multisite cluster.

## Changelog

---

- build/ops: Destruction of basic\_string \_GLIBCXX\_USE\_CXX11\_ABI=0 and C++17 mode results in invalid delete ([issue#38177](#), [pr#26593](#), Kefu Chai, Jason Dillaman)
- build/ops: rpm: require ceph-base instead of ceph-common ([issue#37620](#), [pr#25809](#), Sébastien Han)
- build/ops: run-make-check.sh ccache tweaks ([issue#24817](#), [issue#24777](#), [pr#25153](#), Nathan Cutler, Jonathan Brielmaier, Erwan Velu)
- ceph-create-keys: fix octal notation for Python 3 without losing compatibility with Python 2 ([issue#37641](#), [pr#25531](#), James Page)
- cephfs: MDCache::finish\_snaprealm\_reconnect() create and drop MClientSnap message ([issue#38285](#), [pr#26472](#), "Yan, Zheng")
- cephfs: mgr/status: fix fs status subcommand did not show standby-replay MDS' perf info ([issue#36399](#), [pr#25031](#), Zhi Zhang)
- ceph-objectstore-tool: Dump hashinfo ([issue#37597](#), [pr#25721](#), David Zafman)
- ceph-volume-client: allow setting mode of CephFS volumes ([issue#36651](#), [pr#25413](#),

Tom Barron)

- ceph-volume: enable device discards ([issue#36532](#), [pr#25749](#), Jonas Jelten)
- ceph-volume: fix JSON output in inventory ([issue#37390](#), [pr#25923](#), Sebastian Wagner)
- ceph-volume: Fix TypeError: join() takes exactly one argument (2 given) ([issue#37595](#), [pr#25771](#), Sebastian Wagner)
- ceph-volume normalize comma to dot for string to int conversions ([issue#37442](#), [pr#25775](#), Alfredo Deza)
- ceph-volume: revert partition as disk ([issue#37506](#), [pr#26294](#), Jan Fajerski)
- ceph-volume: set permissions right before prime-osd-dir ([issue#37486](#), [pr#25777](#), Andrew Schoen, Alfredo Deza)
- ceph-volume tests/functional declare ceph-ansible roles instead of importing them ([issue#37805](#), [pr#25837](#), Alfredo Deza)
- ceph-volume zap: improve zapping to remove all partitions and all LVs, encrypted or not ([issue#37449](#), [pr#25351](#), Alfredo Deza)
- cli: dump osd-fsid as part of osd find <id> ([issue#37966](#), [pr#26035](#), Noah Watkins)
- client: do not move f->pos until success write ([issue#37546](#), [pr#25683](#), Junhui Tang)
- client: fix failure in quota size limitation when using samba ([issue#37547](#), [pr#25678](#), Junhui Tang)
- client: fix fuse client hang because its pipe to mds is not ok ([issue#36079](#), [pr#25903](#), Guan yunfei)
- client: retry remount on dcache invalidation failure ([issue#27657](#), [pr#24695](#), Venky Shankar)
- client: session flush does not cause cap release message flush ([issue#38009](#), [pr#26424](#), Patrick Donnelly)
- cmake: do not pass -B{symbolic,symbolic-functions} to linker on FreeBSD ([issue#36717](#), [pr#25525](#), Willem Jan Withagen)
- common: fix memory leaks in WeightedPriorityQueue ([issue#36248](#), [pr#25295](#), Radoslaw Zarzynski)
- common: fix missing include boost/noncopyable.hpp ([issue#38178](#), [pr#26277](#), Willem Jan Withagen)
- core: list-inconsistent-obj output truncated, causing osd-scrub-repair.sh failure ([issue#37653](#), [pr#25603](#), David Zafman)

- core: luminous->(mimic, nautilus): PGMapDigest decode error on luminous end ([issue#38295](#), [pr#26451](#), Sage Weil)
- core: Objecter::calc\_op\_budget: Fix invalid access to extent union member ([issue#37932](#), [pr#26066](#), Simon Ruggier)
- core: scrub warning check incorrectly uses mon scrub interval ([issue#37264](#), [pr#26493](#), David Zafman)
- deep fsck fails on inspecting very large onodes ([issue#38065](#), [pr#26291](#), Igor Fedotov)
- doc: pin the version for “breathe” to 4.1.11 ([issue#38229](#), [pr#26333](#), Alfredo Deza)
- doc: rados/configuration: refresh osdmap section ([issue#38051](#), [pr#26373](#), Ilya Dryomov)
- doc: updated Ceph documentation links ([issue#37793](#), [pr#26180](#), James McClune)
- doc/user-management: Remove obsolete reset caps command ([issue#37663](#), [pr#25607](#), Brad Hubbard)
- journal: max journal order is incorrectly set at 64 ([issue#37541](#), [pr#25957](#), Mykola Golub)
- librbd: fix missing unblock\_writes if shrink is not allowed ([issue#36778](#), [pr#25252](#), runsisi)
- librbd: reset snaps in rbd\_snap\_list() ([issue#37508](#), [pr#25459](#), Kefu Chai)
- mds: broadcast quota message to client when disable quota ([issue#38054](#), [pr#26292](#), Junhui Tang)
- mds: create separate config for heartbeat timeout ([issue#37674](#), [pr#26010](#), Patrick Donnelly)
- mds: directories pinned keep being replicated back and forth between exporting mds and importing mds ([issue#37368](#), [pr#25521](#), Xuehan Xu)
- mds: disallow dumping huge caches to formatter ([issue#36703](#), [pr#25642](#), Venky Shankar)
- mds: do not call Journaler::\_trim twice ([issue#37566](#), [pr#25561](#), Tang Junhui)
- mds: fix bug filelock stuck at LOCK\_XSYN leading client can't read data ([issue#37333](#), [pr#25676](#), Guan yunfei)
- mds: fix incorrect l\_pq\_executing\_ops statistics when meet an invalid item in purge queue ([issue#37567](#), [pr#25559](#), Junhui Tang)
- mds: fix potential re-evaluate stray dentry in \_unlink\_local\_finish ([issue#38263](#),

pr#26474, Zhi Zhang)

- mds: fix races of updating wanted caps ([issue#37464](#), [pr#25680](#), "Yan, Zheng")
- mds: handle fragment notify race ([issue#36035](#), [pr#26252](#), "Yan, Zheng")
- mds: handle state change race ([issue#37594](#), [pr#26051](#), "Yan, Zheng")
- mds: log evicted clients to clog/dbg ([issue#37639](#), [pr#25857](#), Patrick Donnelly)
- MDSMonitor: allow beacons from stopping MDS that was laggy ([issue#37724](#), [pr#25685](#), Patrick Donnelly)
- MDSMonitor: missing osdmon writeable check ([issue#37929](#), [pr#26069](#), Patrick Donnelly)
- mds: purge queue recovery hangs during boot if PQ journal is damaged ([issue#37543](#), [pr#26055](#), Patrick Donnelly)
- mds: PurgeQueue write error handler does not handle EBLACKLISTED ([issue#37394](#), [pr#25523](#), Patrick Donnelly)
- mds: remove duplicated l\_mdc\_num\_strays perfcounter set ([issue#37516](#), [pr#25681](#), Zhi Zhang)
- mds: remove wrong assertion in Locker::snapflush\_nudge ([issue#37721](#), [pr#25885](#), "Yan, Zheng")
- mds: runs out of file descriptors after several respawns ([issue#35850](#), [pr#25822](#), Patrick Donnelly)
- mds: severe internal fragment when decoding xattr\_map from log event ([issue#37399](#), [pr#25519](#), "Yan, Zheng")
- mds: trim cache after journal flush ([issue#38010](#), [pr#26214](#), Patrick Donnelly)
- mds: wait shorter intervals if beacon not sent ([issue#36367](#), [pr#25980](#), Patrick Donnelly)
- mgr: add get\_latest\_counter() to C++ -> Python interface ([issue#38138](#), [pr#26074](#), Jan Fajerski)
- mgr/balancer: add cmd to list all plans ([issue#37418](#), [pr#25293](#), Yang Honggang)
- mgr/balancer: add crush\_compat\_metrics param to change optimization keys ([issue#37412](#), [pr#25291](#), Dan van der Ster)
- mgr/dashboard: Set mirror\_mode to None ([issue#37870](#), [pr#26009](#), Sebastian Wagner)
- mgr: deadlock: \_check\_auth\_rotating possible clock skew, rotating keys expired way too early ([issue#23460](#), [pr#26426](#), Yan Jun)

- mgr: prometheus: added bluestore db and wal devices to ceph\_disk\_occupation metric ([issue#36627](#), [pr#25218](#), Konstantin Shalygin)
- mgr: race between daemon state and service map in 'service status' ([issue#36656](#), [pr#25368](#), Mykola Golub)
- mgr/restful: fix py got exception when get osd info ([issue#38182](#), [pr#26200](#), Boris Ranto, zouaiguo)
- mgr: various python3 fixes ([issue#37415](#), [pr#25292](#), Noah Watkins)
- mgr will refuse connection from the monitor who starts behind it ([issue#37753](#), [pr#26235](#), Xinying Song)
- mgr/zabbix: Send more PG information to Zabbix ([issue#38180](#), [pr#25944](#), Wido den Hollander)
- mon: A PG with PG\_STATE\_REPAIR doesn't mean damaged data, PG\_STATE\_IN... ([issue#38070](#), [pr#26304](#), David Zafman)
- mon: log last command skips latest entry ([issue#36679](#), [pr#25526](#), John Spray)
- mon: mark REMOVE\_SNAPS messages as no\_reply ([issue#37568](#), [pr#25782](#), "Yan, Zheng")
- mon/OSDMonitor: do not populate void pg\_temp into nextmap ([issue#37784](#), [pr#25844](#), Aleksei Zakharov)
- mon: shutdown messenger early to avoid accessing deleted logger ([issue#37780](#), [pr#25846](#), ningtao)
- msg/async: backport recent messenger fixes ([issue#36497](#), [issue#37778](#), [pr#25958](#), xie xingguo)
- msg/async: crashes when authenticator provided by verify\_authorizer not implemented ([issue#36443](#), [pr#25299](#), Sage Weil)
- multisite: es sync null versioned object failed because of olh info ([issue#23842](#), [issue#23841](#), [pr#25578](#), Tianshan Qu, Shang Ding)
- os/bluestore: fixup access a destroy cond cause deadlock or undefine ([issue#37733](#), [pr#26260](#), linbing)
- os/bluestore: KernelDevice::read() does the EIO mapping now ([issue#36455](#), [pr#25854](#), Radoslaw Zarzynski)
- os/bluestore: rename does not old ref to replacement onode at old name ([issue#36541](#), [pr#25313](#), Sage Weil)
- osd: Add support for osd\_delete\_sleep configuration value ([issue#36474](#), [pr#25507](#), Jianpeng Ma, David Zafman)
- osd-backfill-stats.sh fails in rados/standalone/osd.yaml ([issue#37393](#),

[issue#35982](#), [pr#26329](#), Sage Weil, David Zafman)

- osd: backport recent upmap fixes ([issue#37940](#), [issue#37881](#), [pr#26128](#), huangjun, xie xingguo)
- osdc/Objecter: update op\_target\_t::paused in \_calc\_target ([issue#37398](#), [pr#25718](#), Song Shun, runsisi)
- osd: failed assert when osd\_memory\_target options mismatch ([issue#37507](#), [pr#25605](#), xie xingguo)
- osd: force-backfill sets forced\_recovery instead of forced\_backfill in 13.2.1 ([issue#27985](#), [pr#26324](#), xie xingguo)
- osd/mon: fix upgrades for pg log hard limit ([issue#36686](#), [pr#26206](#), Neha Ojha)
- osd/OSDMap: cancel mapping if target osd is out ([issue#37501](#), [pr#25699](#), ningtao, xie xingguo)
- osd/OSD: OSD::mkfs asserts when reusing disk with existing superblock ([issue#37404](#), [pr#25385](#), Igor Fedotov)
- osd/PG.cc: account for missing set irrespective of last\_complete ([issue#37919](#), [pr#26239](#), Neha Ojha)
- osd/PrimaryLogPG: fix the extent length error of the sync read ([issue#37680](#), [pr#25708](#), Xiaofei Cui)
- osd: Prioritize user specified scrubs ([issue#37269](#), [pr#25513](#), David Zafman)
- os/filestore: ceph\_abort() on fsync(2) or fdatasync(2) failure ([issue#38258](#), [pr#26438](#), Sage Weil)
- pybind/mgr: drop unnecessary iterkeys usage to make py-3 compatible ([issue#37581](#), [pr#25759](#), Mykola Golub)
- pybind/mgr/status: fix ceph fs status in py3 environments ([issue#37573](#), [pr#25694](#), Jan Fajerski)
- qa: pjd test appears to require more than 3h timeout for some configurations ([issue#36594](#), [pr#25557](#), Patrick Donnelly)
- qa/rados/upgrade: align thrashing with upgrade suite, don't import/export pgns ([issue#37665](#), [pr#25856](#), Sage Weil)
- qa/tasks/radosbench: default to 64k writes ([issue#37797](#), [pr#26354](#), Sage Weil)
- qa: test\_damage needs to silence MDS\_READ\_ONLY ([issue#37944](#), [pr#26072](#), Patrick Donnelly)
- qa: test\_damage performs truncate test on same object repeatedly ([issue#37836](#), [issue#37837](#), [pr#26047](#), Patrick Donnelly)

- qa: teuthology may hang on diagnostic commands for fuse mount ([issue#36390](#), [pr#25515](#), Patrick Donnelly)
- qa: whitelist cap revoke warning ([issue#25188](#), [pr#26496](#), Patrick Donnelly)
- qa/workunits/rados/test\_health\_warnings: prevent out osds ([issue#37776](#), [pr#25850](#), Sage Weil)
- qa: wrong setting for msgr failures ([issue#36676](#), [pr#25517](#), Patrick Donnelly)
- rbd: fix delay time calculation for trash move ([issue#37861](#), [pr#25954](#), Mykola Golub)
- rgw: debug logging for v4 auth does not sanitize encryption keys ([issue#37847](#), [pr#26003](#), Casey Bodley)
- rgw: Don't treat colons specially in resource part of ARN ([issue#23817](#), [pr#25386](#), Adam C. Emerson)
- rgw: fails to start on Fedora 28 from default configuration ([issue#24228](#), [pr#26129](#), Matt Benjamin)
- rgw: feature - log successful bucket resharding events ([issue#37647](#), [pr#25740](#), J. Eric Ivancich)
- rgw\_file: user info never synced since librgw init ([issue#37527](#), [pr#25485](#), Tao Chen)
- rgw: fix max-size in radosgw-admin and REST Admin API ([issue#37517](#), [pr#25449](#), Nick Erdmann)
- rgw: fix version bucket stats ([issue#21429](#), [pr#25643](#), Shasha Lu)
- rgw: handle S3 version 2 pre-signed urls with meta-data ([issue#23470](#), [pr#25899](#), Matt Benjamin)
- rgw: master zone deletion without a zonegroup rm would break rgw rados init ([issue#37328](#), [pr#25511](#), Abhishek Lekshmanan)
- rgw: multisite: sync gets stuck retrying deletes that fail with ERR\_PRECONDITION\_FAILED ([issue#37448](#), [pr#25505](#), Casey Bodley)
- rgw: Object can still be deleted even if s3>DeleteObject policy is set ([issue#37403](#), [pr#26309](#), Enming.Zhang)
- rgw: "radosgw-admin bucket rm ... -purge-objects" can hang ([issue#38134](#), [pr#26266](#), J. Eric Ivancich)
- rgw: radosgw-admin: translate reshard status codes (trivial) ([issue#36486](#), [pr#25198](#), Matt Benjamin)
- rgw: rgwgc: process coredump in some special case ([issue#23199](#), [pr#25624](#),

zhaokun)

- rpm: Use hardened LDFLAGS ([issue#36316](#), [pr#25171](#), Boris Ranto)

## v13.2.4 Mimic

---

This is the fourth bugfix release of the Mimic v13.2.x long term stable release series. This release includes two security fixes that were tested but inadvertently excluded from the final v13.2.3 release build.

### Changelog

---

- CVE-2018-16846: rgw: enforce bounds on max-keys/max-uploads/max-parts ([issue#35994](#))
- CVE-2018-14662: mon: limit caps allowed to access the config store

## v13.2.3 Mimic

---

This is the third bugfix release of the Mimic v13.2.x long term stable release series. This release contains many fixes across all components of Ceph. We recommend that all users upgrade.

- The default memory utilization for the mons has been increased somewhat. Rocksdb now uses 512 MB of RAM by default, which should be sufficient for small to medium-sized clusters; large clusters should tune this up. Also, the mon\_osd\_cache\_size has been increase from 10 OSDMaps to 500, which will translate to an additional 500 MB to 1 GB of RAM for large clusters, and much less for small clusters.
- Ceph v13.2.2 includes a wrong backport, which may cause mds to go into ‘damaged’ state when upgrading Ceph cluster from previous version. The bug is fixed in v13.2.3. If you are already running v13.2.2, upgrading to v13.2.3 does not require special action.
- The bluestore\_cache\_\* options are no longer needed. They are replaced by osd\_memory\_target, defaulting to 4GB. BlueStore will expand and contract its cache to attempt to stay within this limit. Users upgrading should note this is a higher default than the previous bluestore\_cache\_size of 1GB, so OSDs using BlueStore will use more memory by default. For more details, see the [BlueStore docs](#).
- This version contains an upgrade bug, <http://tracker.ceph.com/issues/36686>, due to which upgrading during recovery/backfill can cause OSDs to fail. This bug can be worked around, either by restarting all the OSDs after the upgrade, or by upgrading when all PGs are in “active+clean” state. If you have already successfully upgraded to 13.2.2, this issue should not impact you. Going forward, we are working on a clean upgrade path for this feature.

# Changelog

---

- build/ops: Can't compile Ceph on Fedora 29 as it doesn't recognize python\*3\*-tox as an install Tox ([issue#18163](#), [issue#37301](#), [issue#37422](#), [pr#25294](#), Nathan Cutler, Brad Hubbard)
- build/ops: debian: correct ceph-common relationship with older radosgw package ([pr#25115](#), Matthew Vernon)
- ceph-bluestore-tool: fix set label functionality for specific keys ([pr#24352](#), Igor Fedotov)
- ceph fs add\_data\_pool applies pool application metadata incorrectly ([issue#36203](#), [issue#36028](#), [pr#24470](#), John Spray)
- cephfs: client: explicitly show blacklisted state via asok status command ([issue#36457](#), [issue#36352](#), [pr#24993](#), Jonathan Brielmeyer, Zhi Zhang)
- cephfs: client: request next osdmap for blacklisted client ([issue#36668](#), [issue#36690](#), [pr#24987](#), Zhi Zhang)
- cephfs-journal-tool: wrong layout info used ([issue#24933](#), [issue#24644](#), [pr#24583](#), Gu Zhongyan)
- cephfs: some tool commands silently operate on only rank 0, even if multiple ranks exist ([issue#36218](#), [pr#25036](#), Venky Shankar)
- ceph-fuse: add to selinux profile ([issue#36103](#), [issue#36197](#), [pr#24439](#), Patrick Donnelly)
- ceph-volume: activate option -auto-detect-objectstore respects -no-systemd ([issue#36249](#), [pr#24357](#), Alfredo Deza)
- ceph-volume add device\_id to inventory listing ([pr#25349](#), Jan Fajerski)
- ceph-volume: add inventory command ([issue#24972](#), [pr#25013](#), Jan Fajerski)
- ceph-volume Additional work on ceph-volume to add some choose\_disk capabilities ([issue#36446](#), [pr#24782](#), Erwan Velu)
- ceph-volume add new ceph-handlers role from ceph-ansible ([issue#36251](#), [pr#24337](#), Alfredo Deza)
- ceph-volume: adds a -prepare flag to lvm batch ([issue#36363](#), [pr#24760](#), Andrew Schoen)
- ceph-volume: allow to specify -cluster-fsid instead of reading from ceph.conf ([issue#26953](#), [pr#25116](#), Alfredo Deza)
- ceph\_volume\_client: py3 compatible ([issue#26850](#), [issue#17230](#), [pr#24443](#), Rishabh Dave, Patrick Donnelly)

- ceph-volume custom cluster names fail on filestore trigger ([issue#27210](#), [pr#24279](#), Alfredo Deza)
- ceph-volume: do not send (lvm) stderr/stdout to the terminal, use the logfile ([issue#36492](#), [pr#24740](#), Alfredo Deza)
- ceph-volume enable --no-systemd flag for simple sub-command ([issue#36470](#), [pr#25011](#), Alfredo Deza)
- ceph-volume: fix journal and filestore data size in lvm batch -report ([issue#36242](#), [pr#24306](#), Andrew Schoen)
- ceph-volume: lsblk can fail to find PARTLABEL, must fallback to blkid ([issue#36098](#), [pr#24334](#), Alfredo Deza)
- ceph-volume lvm.prepare update help to indicate partitions are needed, not devices ([issue#24795](#), [pr#24449](#), Alfredo Deza)
- ceph-volume: make lvm batch idempotent ([pr#24588](#), Andrew Schoen)
- ceph-volume: patch Device when testing ([issue#36768](#), [pr#25066](#), Alfredo Deza)
- ceph-volume: reject devices that have existing GPT headers ([issue#27062](#), [pr#25103](#), Andrew Schoen)
- ceph-volume: remove LVs when using zap -destroy ([pr#25100](#), Alfredo Deza)
- ceph-volume remove version reporting from help menu ([issue#36386](#), [pr#24753](#), Alfredo Deza)
- ceph-volume: rename Device property valid to available ([issue#36701](#), [pr#25133](#), Jan Fajerski)
- ceph-volume: skip processing devices that don't exist when scanning system disks ([issue#36247](#), [pr#24381](#), Alfredo Deza)
- ceph-volume systemd import main so console\_scripts work for executable ([issue#36648](#), [pr#24852](#), Alfredo Deza)
- ceph-volume tests install ceph-ansible's requirements.txt dependencies ([issue#36672](#), [pr#24959](#), Alfredo Deza)
- ceph-volume tests.systemd update imports for systemd module ([issue#36704](#), [pr#24957](#), Alfredo Deza)
- ceph-volume: use console\_scripts ([issue#36601](#), [pr#24838](#), Mehdi Abaakouk)
- ceph-volume util.encryption don't push stderr to terminal ([issue#36246](#), [pr#24826](#), Alfredo Deza)
- ceph-volume util.encryption robust blkid+lsblk detection of lockbox ([pr#24980](#), Alfredo Deza)

- client: fix use-after-free in Client::link() ([issue#35841](#), [issue#24557](#), [pr#24187](#), "Yan, Zheng")
- client: statfs inode count odd ([issue#35940](#), [issue#24849](#), [pr#24377](#), Rishabh Dave)
- client:two ceph-fuse client, one can not list out files created by another ([issue#27051](#), [issue#35934](#), [pr#24295](#), Peng Xie)
- client: update ctime when modifying file content ([issue#35945](#), [issue#36134](#), [pr#24385](#), "Yan, Zheng")
- common: get real hostname from container/pod environment ([pr#23916](#), Sage Weil)
- core: \_aio\_log\_start inflight overlap of 0x10000~1000 with [65536~4096] ([issue#36754](#), [issue#36625](#), [pr#25062](#), Jonathan Brielmair, Yang Honggang)
- core: FAILED assert(osdmap\_manifest.pinned.empty()) in OSDMonitor::prune\_init() ([issue#24612](#), [issue#35071](#), [pr#24918](#), Joao Eduardo Luis)
- core: Interactive mode CLI prints no output since Mimic ([issue#36358](#), [issue#36432](#), [pr#24971](#), John Spray, Mohamad Gebai)
- core: mgr crash on scrub of unconnected osd ([issue#36110](#), [issue#36465](#), [pr#25029](#), Sage Weil)
- core: mon osdmap cash too small during upgrade to mimic ([issue#36505](#), [pr#25019](#), Sage Weil)
- core: monstore tool rebuild does not generate creating\_pgs ([issue#36306](#), [issue#36433](#), [pr#25016](#), Sage Weil)
- core: Objecter: add ignore cache flag if got redirect reply ([issue#36658](#), [pr#25075](#), Iain Buclaw, Jonathan Brielmair)
- core: objecter cannot resend split-dropped op when racing with con reset ([issue#22544](#), [issue#35843](#), [pr#24970](#), Sage Weil)
- core: os/bluestore: cache autotuning and memory limit ([issue#37340](#), [pr#25283](#), Josh Durgin, Mark Nelson)
- core: rados rm -force-full is blocked when cluster is in full status ([issue#36435](#), [pr#25017](#), Yang Honggang)
- crush/CrushWrapper: fix crush tree json dumper ([issue#36150](#), [pr#24481](#), Oshyn Song)
- debian/control: require fuse for ceph-fuse ([issue#21057](#), [pr#24037](#), Thomas Serlin)
- doc: add ceph-volume inventory sections ([pr#25130](#), Jan Fajerski)
- doc: fix broken fstab url in cephfs/fuse ([issue#36286](#), [issue#36313](#), [pr#24441](#), Jos Collin)

- doc: Put command template into literal block ([pr#25000](#), Alexey Stupnikov)
- doc: remove deprecated ‘scrubq’ from ceph(8) ([issue#35813](#), [issue#35855](#), [pr#24210](#), Ruben Kerkhof)
- docs: backport edit on github changes ([pr#25362](#), Neha Ojha, Noah Watkins)
- doc: Typo error on cephfs/fuse/ ([issue#36180](#), [issue#36308](#), [pr#24420](#), Karun Josy)
- ec: src/common/interval\_map.h: 161: FAILED assert(len > 0) ([issue#21931](#), [issue#22330](#), [pr#24581](#), Neha Ojha)
- fsck: cid is improperly matched to oid ([issue#36146](#), [issue#36551](#), [issue#36099](#), [issue#32731](#), [pr#24480](#), Kefu Chai, Sage Weil)
- kernel\_untar\_build.sh: bison: command not found ([issue#36121](#), [pr#24241](#), Neha Ojha)
- libcephfs: expose CEPH\_SETATTR\_MTIME\_NOW and CEPH\_SETATTR\_ATIME\_NOW ([issue#36205](#), [issue#35961](#), [pr#24464](#), Zhu Shangzhong)
- librados application’s symbol could conflict with the libceph-common ([issue#26839](#), [issue#25154](#), [pr#24708](#), Kefu Chai)
- librbd: blacklisted client might not notice it lost the lock ([issue#34534](#), [pr#24401](#), Jason Dillaman)
- librbd: ensure exclusive lock acquired when removing sync point snaps... ([issue#35714](#), [issue#24898](#), [pr#24137](#), Mykola Golub)
- librbd: fixed assert when flattening clone with zero overlap ([issue#35957](#), [issue#35702](#), [pr#24356](#), Jason Dillaman)
- librbd: journaling unable request can not be sent to remote lock owner ([issue#26939](#), [issue#35712](#), [pr#24122](#), Mykola Golub)
- librbd: object map improperly flagged as invalidated ([issue#24516](#), [issue#36225](#), [pr#24413](#), Jason Dillaman)
- librgw: crashes in multisite configuration ([issue#36302](#), [issue#36415](#), [pr#24908](#), Casey Bodley)
- mds: allows client to create .. and . dirents ([issue#32104](#), [pr#24384](#), Venky Shankar)
- mds: curate priority of perf counters sent to mgr ([issue#35938](#), [issue#26991](#), [issue#32090](#), [issue#35837](#), [pr#24467](#), Patrick Donnelly, Venky Shankar)
- mds: evict cap revoke non-responding clients ([pr#24661](#), Venky Shankar)
- mimic:mds: fix mds damaged due to unexpected journal length ([issue#36199](#), [pr#24463](#), Zhi Zhang)

- mds: internal op missing events time ‘throttled’, ‘all\_read’, ‘dispatched’ ([issue#36114](#), [issue#36195](#), [pr#24411](#), Yanhu Cao)
- mds: migrate strays part by part when shutdown mds ([issue#26926](#), [issue#32092](#), [pr#24435](#), “Yan, Zheng”)
- mds: optimize the way how max export size is enforced ([issue#25131](#), [pr#23952](#), “Yan, Zheng”)
- mds: print is\_laggy message once ([issue#35250](#), [issue#35719](#), [pr#24161](#), Patrick Donnelly)
- mds: rctime may go back ([issue#35916](#), [issue#36136](#), [pr#24379](#), “Yan, Zheng”)
- mds: rctime not set on system inode (root) at startup ([issue#36221](#), [issue#36461](#), [pr#25042](#), Patrick Donnelly)
- mds: reset heartbeat map at potential time-consuming places ([issue#26858](#), [pr#23506](#), Yan, Zheng, “Yan, Zheng”)
- mds: src/mds/MDLog.cc: 281: FAILED ceph\_assert(!capped) during max\_mds thrashing ([issue#36350](#), [issue#37093](#), [pr#25095](#), “Yan, Zheng”, Jonathan Brielmairer)
- mgr/DaemonServer: fix Session leak ([pr#24233](#), Sage Weil)
- mgr/dashboard: Add http support to dashboard ([issue#36069](#), [pr#24734](#), Boris Ranto, Wido den Hollander)
- mgr/dashboard: Add support for URI encode ([issue#24621](#), [issue#26856](#), [issue#24907](#), [pr#24488](#), Tiago Melo)
- mgr/dashboard: Progress bar does not stop in TableKeyValueComponent ([issue#35925](#), [pr#24258](#), Volker Theile)
- mgr/dashboard: Remove fieldsets when using CdTable ([issue#27851](#), [issue#26999](#), [pr#24478](#), Tiago Melo)
- mgr: hold lock while accessing the request list and submittin request ([pr#25113](#), Jerry Lee)
- mgr: [restful] deep\_scrub is not a valid OSD command ([issue#36720](#), [issue#36749](#), [pr#25040](#), Boris Ranto)
- mon: mgr options not parse properly ([issue#35076](#), [issue#35836](#), [pr#24176](#), Sage Weil)
- mon/OSDMonitor: invalidate max\_failed\_since on cancel\_report ([issue#35930](#), [issue#35860](#), [pr#24281](#), xie xingguo)
- mon: test if gid exists in pending for prepare\_beacon ([issue#35848](#), [pr#24272](#), Patrick Donnelly)

- msg/async: clean up local buffers on dispatch ([issue#36127](#), [issue#35987](#), [pr#24386](#), Greg Farnum)
- msg: ceph\_abort() when there are enough accepter errors in msg server ([issue#36219](#), [pr#25045](#), [penglaixy@gmail.com](#))
- msg: challenging authorizer messages appear at debug\_ms=0 ([issue#35251](#), [issue#35717](#), [pr#24113](#), Patrick Donnelly)
- multisite: data full sync does not limit concurrent bucket sync ([issue#26897](#), [issue#36216](#), [pr#24536](#), Casey Bodley)
- multisite: data sync error repo processing does not back off on empty ([issue#35979](#), [issue#26938](#), [pr#24319](#), Casey Bodley)
- multisite: incremental data sync makes unnecessary call to RGWReadRemoteDataLogShardInfoCR ([issue#35977](#), [issue#26952](#), [pr#24710](#), Casey Bodley)
- multisite: intermittent test\_bucket\_index\_log\_trim failures ([issue#36201](#), [issue#36034](#), [pr#24400](#), Casey Bodley)
- multisite: invalid read in RGWCloneMetaLogCoroutine ([issue#36208](#), [issue#35851](#), [pr#24414](#), Casey Bodley)
- multisite: segfault on shutdown/realm reload ([issue#35857](#), [issue#35543](#), [pr#24235](#), Casey Bodley)
- os/bluestore: fix bloom filter num entry miscalculation in repairer ([issue#25001](#), [pr#24339](#), Igor Fedotov)
- os/bluestore: handle spurious read errors ([issue#22464](#), [pr#24647](#), Paul Emmerich)
- osd: add creating to pg\_string\_state ([issue#36174](#), [issue#36298](#), [pr#24601](#), Dan van der Ster)
- osd: backport recent upmap fixes ([pr#25419](#), ningtao, xie xingguo)
- osdc/Objecter: possible race condition with connection reset ([issue#36183](#), [issue#36296](#), [pr#24600](#), Jason Dillaman)
- osd: crash in OpTracker::unregister\_inflight\_op via OSD::get\_health\_metrics ([issue#24889](#), [pr#23026](#), Radoslaw Zarzynski)
- osdc: reduce ObjectCacher's memory fragments ([issue#36192](#), [issue#36643](#), [pr#24873](#), "Yan, Zheng")
- osd/ECBackend: don't get result code of subchunk-read overwritten ([issue#35959](#), [issue#21769](#), [pr#24298](#), songweibin)
- OSDMapMapping does not handle active.size() > pool size ([issue#26866](#),

- issue#35936, pr#24431, Sage Weil)
- osd/PG: avoid choose\_acting picking want with > pool size items (issue#35963, issue#35924, pr#24344, Sage Weil)
- osd/PrimaryLogPG: fix potential pg-log overtrimming (pr#24309, xie xingguo)
- osd: race condition opening heartbeat connection (issue#36637, issue#36602, pr#25026, Sage Weil)
- osd: RBD client IOPS pool stats are incorrect (2x higher; includes IO hints as an op) (issue#24909, issue#36557, pr#25024, Jason Dillaman)
- osd: Remove old bft= which has been superceded by backfill (issue#36292, issue#36170, pr#24573, David Zafman)
- qa: add test that builds example librados programs (issue#36228, issue#15100, pr#24537, Nathan Cutler)
- qa/ceph-ansible: Specify stable-3.2 branch (pr#25191, Brad Hubbard)
- qa: extend timeout for SessionMap flush (issue#36156, pr#24438, Patrick Donnelly)
- qa: fsstress workunit does not execute in parallel on same host without clobbering files (issue#36278, issue#24177, issue#36323, issue#36184, issue#36165, issue#36153, pr#24408, Patrick Donnelly)
- qa: increase rm timeout for workunit cleanup (issue#36501, issue#36365, pr#24684, Patrick Donnelly)
- qa: install dependencies for rbd\_workunit\_kernel\_untar\_build (issue#35074, issue#35077, pr#24240, Ilya Dryomov)
- qa: remove knfs site from future releases (issue#36075, issue#36102, pr#24269, Yuri Weinstein)
- qa/suites/rados/thrash-old-clients: exclude packages for hammer, jewel (pr#25193, Neha Ojha)
- qa/suites/rgw/verify/tasks/cls\_rgw: test cls\_rgw (issue#25024, pr#23197, Casey Bodley, Sage Weil)
- qa/tasks/qemu: use unique clone directory to avoid race with workunit (issue#36542, issue#36569, pr#24811, Jason Dillaman)
- qa: test\_recovery\_pool tries asok on wrong node (issue#24928, issue#24858, pr#23087, Patrick Donnelly)
- qa: tolerate failed rank while waiting for state (issue#36280, issue#35828, pr#24572, Patrick Donnelly)
- qa/workunits: replace 'realpath' with 'readlink -f' in fsstress.sh (issue#36409,

- issue#36430, issue#35538, pr#24622, Ilya Dryomov, Jason Dillaman)
- RADOS: probably missing clone location for async\_recovery\_targets (issue#35964, issue#35546, pr#24345, xie xingguo)
  - mimic:rbd: fix error import when the input is a pipe (issue#35705, issue#34536, pr#24002, songweibin)
  - [rbd-mirror] failed assertion when updating mirror status (issue#36084, issue#36120, pr#24321, Jason Dillaman)
  - rbd: [rbd-mirror] forced promotion after killing remote cluster results in stuck state (issue#36659, issue#36693, pr#24952, Jonathan Brielmair, Jason Dillaman)
  - rbd: [rbd-mirror] periodic mirror status timer might fail to be scheduled (issue#36500, issue#36555, pr#24916, Jason Dillaman)
  - rbd: rbd-nbd: do not ceph\_abort() after print the usages (issue#36660, issue#36713, pr#24988, Shiyang Ruan)
  - rbd: TokenBucketThrottle: use reference to m\_blockers.front() and then update it (issue#36529, issue#36475, pr#24915, Dongsheng Yang)
  - Revert "mimic: cephfs-journal-tool: enable purge\_queue journal's event commands" (issue#36346, issue#24604, pr#24485, Xuehan Xu, "Yan, Zheng")
  - rgw: abort\_bucket\_multiparts() ignores individual NoSuchUpload errors (issue#36129, issue#35986, pr#24388, Casey Bodley)
  - rgw-admin: reshards add can add a non existant bucket (issue#36449, issue#36756, pr#25087, Jonathan Brielmair, Abhishek Lekshmanan)
  - rgw: async sync\_object and remove\_object does not access coroutine me... (issue#36138, issue#35905, pr#24417, Tianshan Qu)
  - rgw/beast: drop privileges after binding ports (issue#36041, pr#24436, Paul Emmerich)
  - rgw: beast frontend fails to parse ipv6 endpoints (issue#36662, issue#36734, pr#25079, Jonathan Brielmair, Casey Bodley)
  - rgw: cls\_user\_remove\_bucket does not write the modified cls\_user\_stats (issue#36496, issue#36533, pr#24910, Casey Bodley)
  - rgw: default quota not set in radosgw for Openstack users (issue#24595, issue#36223, pr#24907, Casey Bodley)
  - mimic:rgw: fix chunked-encoding for chunks >1MiB (issue#36125, issue#35990, pr#24363, Robin H. Johnson)
  - rgw: fix deadlock on RGWIndexCompletionManager::stop (issue#26949, issue#35710,

- pr#24101, Yao Zongyou)
- mimic:rgw: fix leak of curl handle on shutdown ([issue#35715](#), [issue#36213](#), [pr#24518](#), Casey Bodley)
  - mimic:rgw: list bucket can not show the object uploaded by RGWPostObj when enable bucket versioning ([pr#24571](#), yuliyang)
  - rgw: radosgw-admin user stats are incorrect when dynamic re-sharding is enabled ([issue#36535](#), [pr#24911](#), Casey Bodley)
  - rgw: raise debug level on redundant data sync error messages ([issue#35830](#), [issue#36140](#), [pr#24418](#), Casey Bodley)
  - rgw: raise default rgw\_curl\_low\_speed\_time to 300 seconds ([issue#35708](#), [issue#27989](#), [pr#24071](#), Casey Bodley)
  - rgw: renew resharding locks to prevent expiration ([issue#36687](#), [issue#27219](#), [issue#34307](#), [pr#24899](#), Orit Wasserman, J. Eric Ivancich)
  - rgw: resharding produces invalid values of bucket stats ([issue#36290](#), [issue#36381](#), [pr#24526](#), Abhishek Lekshmanan)
  - mimic:rgw: return x-amz-version-id: null when delete obj in versioning ([issue#35814](#), [pr#24189](#), yuliyang)
  - rgw: RGWAsyncGetBucketInstanceInfo does not access coroutine memory ([issue#36211](#), [issue#35812](#), [pr#24516](#), Casey Bodley)
  - rgw: set default objecter\_inflight\_ops = 24576 ([issue#36571](#), [issue#25109](#), [pr#24860](#), Jonathan Brielmair, Matt Benjamin)
  - rgw: support server-side encryption when SSL is terminated in a proxy ([issue#36645](#), [issue#27221](#), [pr#24931](#), Jonathan Brielmair, Casey Bodley)
  - rgw: use-after-free from RGWRadosGetOmapKeysCR::~RGWRadosGetOmapKeysCR ([issue#21154](#), [issue#36537](#), [issue#36539](#), [pr#24912](#), Casey Bodley, Sage Weil)
  - rpm: use updated gperftools ([issue#36508](#), [issue#35969](#), [pr#24260](#), Brad Hubbard, Kefu Chai)
  - segv in BlueStore::OldExtent::create ([issue#36592](#), [issue#36526](#), [pr#24745](#), Sage Weil)
  - test/librbd: not valid to have different parents between image snapshots ([issue#36117](#), [pr#24244](#), Jason Dillaman)
  - [test] periodic seg faults within unittest\_librbd ([issue#36220](#), [issue#36238](#), [pr#24711](#), Jason Dillaman)
  - test/rbd\_mirror: race in WaitingOnLeaderReleaseLeader ([issue#36236](#), [issue#36276](#),

[pr#24551](#), Mykola Golub)

- tests: ceph-admin-commands.sh workunit does not log what it's doing ([issue#37153](#), [issue#37089](#), [pr#25085](#), Nathan Cutler)
- tests: librados api aio tests race condition ([issue#24587](#), [issue#36647](#), [pr#25027](#), Josh Durgin)
- tests: make readable.sh fail if it doesn't run anything ([pr#25050](#), Greg Farnum)
- tests: rbd: move OpenStack devstack test to rocky release ([issue#36410](#), [issue#36428](#), [pr#24913](#), Jason Dillaman)
- tests: unittest\_rbd\_mirror: TestMockImageMap.AddInstancePingPongImageTest: Value of: it != peer\_ack\_ctxs->end() ([issue#36683](#), [issue#36689](#), [pr#24946](#), Mykola Golub, Jonathan Brielmaier)
- tests: use timeout for fs asok operations ([issue#36335](#), [issue#36503](#), [pr#25332](#), Patrick Donnelly)
- tests: /usr/bin/ld: cannot find -lradospp in rados mimic ([issue#37396](#), [pr#25285](#), Nathan Cutler)
- test: Use a grep pattern that works across releases ([issue#35845](#), [issue#35909](#), [pr#24017](#), David Zafman)
- tools: ceph-objectstore-tool: Allow target level as first positional ... ([issue#35846](#), [issue#35992](#), [pr#24116](#), David Zafman)

## v13.2.2 Mimic

---

This is the second bugfix release of the Mimic v13.2.x long term stable release series. This release contains many fixes across all components of Ceph. We recommend that all users upgrade.

- This version contains an upgrade bug, <http://tracker.ceph.com/issues/36686>, due to which upgrading during recovery/backfill can cause OSDs to fail. This bug can be worked around, either by restarting all the OSDs after the upgrade, or by upgrading when all PGs are in “active+clean” state.

If you have successfully upgraded to 13.2.2, this issue should not impact you. Going forward, we are working on a clean upgrade path for this feature.

## Changelog

---

- build/ops: Boost system library is no longer required to compile and link example librados program ([issue#25073](#), [issue#25054](#), [pr#23201](#), Nathan Cutler)

- build/ops: debian/rules: fix ceph-mgr .pyc files left behind ([issue#27059](#), [issue#26883](#), [pr#23831](#), Dan Mick)
- build/ops: mimic 13.2.0 doesn't build in Fedora rawhide ([issue#24449](#), [issue#24905](#), [pr#23885](#), Kefu Chai)
- ceph-disk: compatibility fix for python 3 ([pr#24008](#), Tim Serong)
- ceph-disk: return a list instead of an iterator ([pr#23392](#), Alexander Graul)
- cephfs-journal-tool: enable purge\_queue journal's event commands ([issue#24604](#), [issue#26989](#), [pr#23818](#), Xuehan Xu)
- ceph tell osd.x bench writes resulting JSON to stderr instead of stdout ([issue#35942](#), [issue#24022](#), [pr#24041](#), Коренберг Марк, John Spray, Kefu Chai)
- ceph-volume add a \_\_release\_\_ string, to help version-conditional calls ([issue#25169](#), [pr#23333](#), Alfredo Deza)
- ceph-volume: adds test for ceph-volume lvm list /dev/sda ([issue#24784](#), [issue#24957](#), [pr#23349](#), Andrew Schoen)
- ceph-volume: an OSD ID must exist and be destroyed before reuse ([pr#23101](#), Andrew Schoen, Ron Allred)
- ceph-volume: batch: allow journal+block.db sizing on the CLI ([issue#36088](#), [pr#24208](#), Alfredo Deza)
- ceph-volume batch: allow -osds-per-device, default it to 1 ([issue#35913](#), [pr#24079](#), Alfredo Deza)
- ceph-volume batch carve out lvs for bluestore ([issue#34535](#), [pr#24074](#), Alfredo Deza)
- ceph-volume batch command ([pr#23777](#), Alfredo Deza)
- ceph-volume: batch tests for mixed-type of devices ([issue#35535](#), [issue#27210](#), [pr#23966](#), Alfredo Deza)
- ceph\_volume\_client: allow atomic update of RADOS objects ([issue#24173](#), [issue#24863](#), [pr#23878](#), Rishabh Dave)
- CephVolumeClient: delay required after adding data pool to MDSMap ([issue#25206](#), [pr#23725](#), Patrick Donnelly)
- ceph-volume: do not use stdin in luminous ([issue#25173](#), [pr#23368](#), Alfredo Deza)
- ceph-volume: earlier detection for -journal and -filestore flag requirements ([issue#24794](#), [pr#24205](#), Alfredo Deza)
- ceph-volume enable the ceph-osd during lvm activation ([issue#24152](#), [pr#23393](#), Dan van der Ster, Alfredo Deza)

- ceph-volume expand auto engine for multiple devices on filestore ([pr#23807](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: expand auto engine for single type devices on filestore ([pr#23786](#), Alfredo Deza)
- ceph-volume fix zap not working with LVs ([issue#35970](#), [pr#24081](#), Alfredo Deza)
- ceph-volume lvm.activate conditional mon-config on prime-osd-dir ([issue#25216](#), [pr#23400](#), Alfredo Deza)
- ceph-volume: lvm batch allow extra flags (like dmcrypt) for bluestore ([pr#23780](#), Alfredo Deza)
- ceph-volume: lvm batch documentation and man page updates ([pr#23756](#), Alfredo Deza)
- ceph-volume lvm.batch remove non-existent sys\_api property ([issue#34310](#), [pr#23810](#), Alfredo Deza)
- ceph-volume lvm.listing only include devices if they exist ([issue#24952](#), [pr#23149](#), Alfredo Deza)
- ceph-volume: process.call with stdin in Python 3 fix ([issue#24993](#), [pr#23239](#), Alfredo Deza)
- ceph-volume: PVolumes.get() should return one PV when using name or uuid ([issue#24784](#), [pr#23327](#), Andrew Schoen)
- ceph-volume: refuse to zap mapper devices ([issue#24504](#), [pr#22965](#), Andrew Schoen)
- ceph-volume: Restore SELinux context ([pr#23295](#), Boris Ranto)
- ceph-volume: run tests without waiting on ceph repos ([pr#23806](#), Andrew Schoen)
- ceph-volume tests/functional add mgrs daemons to lvm tests ([pr#23784](#), Alfredo Deza)
- ceph-volume: tests.functional inherit SSH\_ARGS from ansible ([pr#23812](#), Alfredo Deza)
- ceph-volume: update batch documentation to explain filestore strategies ([issue#34309](#), [pr#23826](#), Alfredo Deza)
- ceph-volume: update version of ansible to 2.6.x for simple tests ([pr#23269](#), Andrew Schoen)
- client: add inst to asok status output ([issue#24724](#), [issue#24931](#), [pr#23109](#), Patrick Donnelly)
- client: check for unmounted condition before printing debug output ([issue#25213](#), [issue#26914](#), [pr#23603](#), Jeff Layton)

- client: requests that do name lookup may be sent to wrong mds ([issue#26984](#), [issue#26860](#), [pr#23700](#), "Yan, Zheng")
- cls/rgw: add rgw\_usage\_log\_entry type to ceph-dencoder ([issue#35070](#), [pr#23857](#), Vaibhav Bhembre)
- common: check completion condition before waiting ([issue#25007](#), [issue#25222](#), [pr#23435](#), Patrick Donnelly)
- core: deep scrub cannot find the bitrot if the object is cached ([issue#35068](#), [pr#23873](#), Adam C. Emerson, Xiaoguang Wang)
- core: Fix 25085 and 24949 ([pr#23272](#), David Zafman)
- core: force-create-pg broken ([issue#34532](#), [issue#26940](#), [pr#23872](#), Sage Weil)
- core: Limit pg log length during recovery/backfill so that we don't run out of memory ([issue#21416](#), [pr#23403](#), Neha Ojha)
- doc: broken bash example in bluestore migration ([issue#35078](#), [pr#23854](#), Alfredo Deza)
- doc: Fix broken urls ([issue#25185](#), [issue#26916](#), [pr#23607](#), Jos Collin)
- doc: <http://docs.ceph.com/docs/mimic/rados/operations/pg-states/> ([issue#25055](#), [pr#23163](#), Jan Fajerski, Nathan Cutler)
- docs: radosgw: ldap-auth: fixed option name 'rgw\_ldap\_searchfilter' ([issue#32129](#), [pr#23956](#), Konstantin Shalygin)
- filestore: add pgid in filestore pg dir split log message ([issue#25225](#), [pr#23453](#), Vikhyat Umrao)
- kv: MergeOperator name() returns string, and caller calls c\_str() on the temporary ([issue#26907](#), [issue#26875](#), [pr#23865](#), Sage Weil)
- libradosstriper conditional compile ([issue#27213](#), [pr#23869](#), Kefu Chai, Jesse Williamson)
- librbd: deep-copy should not write to objects that cannot exist ([issue#25000](#), [issue#25083](#), [pr#23358](#), Jason Dillaman)
- librbd: validate data pool for self-managed snapshot support ([issue#24945](#), [pr#23560](#), Mykola Golub)
- link against libstdc++ statically ([issue#26880](#), [issue#25209](#), [pr#23490](#), Kefu Chai)
- mds: avoid using g\_conf->get\_val<...>(...) in hot path ([issue#24820](#), [pr#23407](#), "Yan, Zheng")
- mds: calculate load by checking self CPU usage ([issue#26834](#), [issue#26888](#), [pr#23503](#), "Yan, Zheng")

- mds: crash when dumping ops in flight ([issue#26894](#), [issue#26982](#), [pr#23672](#), "Yan, Zheng")
- mds: dump recent events on respawn ([issue#25040](#), [pr#23275](#), Patrick Donnelly)
- mds: explain delayed client\_request due to subtree migration ([issue#26988](#), [issue#24840](#), [pr#23792](#), Yan, Zheng, "Yan, Zheng")
- mds: handle discontinuous mdsmap ([issue#24856](#), [pr#23180](#), "Yan, Zheng")
- mds: health warning for slow metadata IO ([issue#24879](#), [issue#25045](#), [pr#23343](#), "Yan, Zheng")
- mds: increase debug level for dropped client cap msg ([issue#25042](#), [pr#23309](#), Patrick Donnelly)
- mds: introduce cephfs' own feature bits ([issue#14456](#), [issue#24914](#), [pr#23105](#), Yan, Zheng, "Yan, Zheng", Patrick Donnelly)
- mds: mark beacons as high priority ([issue#26905](#), [issue#26899](#), [pr#23565](#), Patrick Donnelly)
- mds: MDBalancer::try\_rebalance() may stop prematurely ([issue#32086](#), [issue#26973](#), [pr#23883](#), "Yan, Zheng")
- MDSMonitor: note ignored beacons/map changes at higher debug level ([issue#26898](#), [issue#26929](#), [pr#23704](#), Patrick Donnelly)
- mds,osd,mon,msg: use intrusive\_ptr for holding Connection::priv ([issue#20924](#), [pr#22339](#), "Yan, Zheng", Kefu Chai)
- mds: print mdsmap processed at low debug level ([issue#25035](#), [pr#23196](#), Patrick Donnelly)
- mds: scrub doesn't always return JSON results ([issue#23958](#), [issue#25037](#), [pr#23225](#), Venky Shankar)
- mds: use fast dispatch to handle MDSBeacon ([issue#23519](#), [issue#26923](#), [pr#23703](#), "Yan, Zheng")
- mgr balancer does not save optimized plan but latest ([issue#32082](#), [issue#27000](#), [pr#23782](#), Stefan Priebe)
- mgr: "balancer execute" only requires read permissions ([issue#26912](#), [issue#25345](#), [pr#23583](#), John Spray)
- mgrc: enable disabling stats via mgr\_stats\_threshold ([issue#25197](#), [issue#26837](#), [pr#23463](#), John Spray)
- mgr/dashboard: Display RGW user/bucket quota max size in human readable form ([issue#35706](#), [pr#24047](#), Volker Theile)

- mgr/dashboard: Escape regex pattern in DeletionModalComponent ([issue#24902](#), [issue#26920](#), [pr#23669](#), Tiago Melo)
- mgr/dashboard: Prevent RGW API user deletion ([pr#22670](#), Volker Theile)
- mgr/dashboard: RestClient can't handle ProtocolError exceptions ([pr#23875](#), Volker Theile)
- mgr/dashboard: RGW is not working if an URL prefix is defined ([pr#23203](#), Volker Theile)
- mgr/dashboard: URL prefix is not working ([issue#25120](#), [pr#23874](#), Ricardo Marques)
- mgr: Ignore daemon if no metadata was returned ([pr#23356](#), Wido den Hollander)
- mgr/MgrClient: Protect daemon\_health\_metrics ([issue#23352](#), [pr#23458](#), Kjetil Joergensen, Brad Hubbard)
- mgr: Sync the prometheus module ([pr#23215](#), Boris Ranto)
- mon: add purge-new ([pr#23259](#), Sage Weil)
- mon: Automatically set expected\_num\_objects for new pools with >=100 PGs per OSD ([issue#24687](#), [issue#25144](#), [pr#23860](#), Douglas Fuller)
- multisite: intermittent failures in test\_bucket\_sync\_disable\_enable ([issue#26895](#), [issue#26980](#), [pr#23856](#), Casey Bodley)
- multisite: object metadata operations are skipped by sync ([issue#24367](#), [issue#24986](#), [pr#23172](#), Casey Bodley)
- object errors found in be\_select\_auth\_object() aren't logged the same ([issue#32108](#), [issue#25108](#), [pr#23870](#), David Zafman)
- os/bluestore: bluestore\_buffer\_hit\_bytes perf counter doesn't reset ([pr#23772](#), Igor Fedotov)
- os/bluestore/BlueStore.cc: 1025: FAILED assert(buffer\_bytes >= b->length) from ObjectStore/StoreTest.ColSplitTest2/2 ([issue#24439](#), [issue#26944](#), [pr#23748](#), Sage Weil)
- os/bluestore: fix assertion in StupidAllocator::get\_fragmentation ([pr#23676](#), Igor Fedotov)
- osd: do\_sparse\_read(): Verify checksum earlier so we will try to repair ([issue#24875](#), [pr#23378](#), David Zafman)
- osd,mon: increase mon\_max\_pg\_per\_osd to 300 ([issue#25176](#), [pr#23861](#), Neha Ojha)
- osd/OSDMap: CRUSH\_TUNABLES5 added in jewel, not kraken ([issue#25057](#), [issue#25101](#), [pr#23226](#), Sage Weil)

- osd/PrimaryLogPG: avoid dereferencing invalid complete\_to ([pr#23951](#), xie xingguo)
- osd: segv in OSDMap::calc\_pg\_upmaps from balancer ([issue#22056](#), [issue#26933](#), [pr#23888](#), Brad Hubbard)
- qa: cfuse\_workunit\_kernel\_untar\_build fails on Ubuntu 18.04 ([issue#26956](#), [issue#26967](#), [issue#24679](#), [pr#23769](#), Patrick Donnelly)
- qa: fix ceph-disk suite and add coverage for ceph-detect-init ([pr#23337](#), Nathan Cutler)
- qa/rgw: patch keystone requirements.txt ([issue#26946](#), [issue#23659](#), [pr#23771](#), Casey Bodley)
- qa/suites/rados: move valgrind test to singleton-flat ([issue#24992](#), [pr#23744](#), Sage Weil)
- qa/tasks: s3a fix mirror ([pr#24038](#), Vasu Kulkarni)
- qa/tests: added OBJECT\_MISPLACED to the whitelist ([pr#23301](#), Yuri Weinstein)
- qa/tests: added v13.2.1 to the mix ([pr#23218](#), Yuri Weinstein)
- qa/tests: update ansible version to 2.5 ([pr#24091](#), Yuri Weinstein)
- rados: not all exceptions accept keyargs ([issue#25178](#), [issue#24033](#), [pr#23335](#), Rishabh Dave)
- rados python bindings use prval from stack ([issue#25204](#), [issue#25175](#), [pr#23863](#), Sage Weil)
- rbd: improved trash snapshot namespace handling ([issue#25121](#), [issue#23398](#), [issue#25114](#), [pr#23559](#), Mykola Golub, Jason Dillaman)
- rgw: add curl\_low\_speed\_limit and curl\_low\_speed\_time config to avoid ([issue#25021](#), [pr#23173](#), Mark Kogan, Zhang Shaowen)
- rgw: change default rgw\_thread\_pool\_size to 512 ([issue#25214](#), [issue#25088](#), [issue#25218](#), [issue#24544](#), [pr#23383](#), Douglas Fuller, Casey Bodley)
- rgw: civetweb fails on urls with control characters ([issue#26849](#), [issue#24158](#), [pr#23855](#), Abhishek Lekshmanan)
- rgw: civetweb: use poll instead of select while waiting on sockets ([issue#35954](#), [pr#24058](#), Abhishek Lekshmanan)
- rgw: do not ignore EEXIST in RGWPutObj::execute ([issue#25078](#), [issue#22790](#), [pr#23206](#), Matt Benjamin)
- rgw: fail to recover index from crash mimic backport ([issue#24640](#), [issue#24629](#), [issue#24280](#), [pr#23118](#), Tianshan Qu)

- rgw\_file: deep stat handling ([issue#26842](#), [issue#24915](#), [pr#23498](#), Matt Benjamin)
- rgw: Fix log level of gc\_iterate\_entries ([issue#23801](#), [issue#26921](#), [pr#23686](#), iliul)
- rgw: Limit the number of lifecycle rules on one bucket ([issue#26845](#), [issue#24572](#), [pr#23521](#), Zhang Shaowen)
- rgw: radosgw-admin: ‘sync error trim’ loops until complete ([issue#24873](#), [issue#24984](#), [pr#23140](#), Casey Bodley)
- rgw: The delete markers generated by object expiration should have owner ([issue#24568](#), [issue#26847](#), [pr#23541](#), Zhang Shaowen)
- rpm: should change ceph-mgr package depency from py-bcrypt to python2-bcrypt ([issue#27212](#), [pr#23868](#), Konstantin Sakhinov)
- rpm: silence osd block chown ([issue#25152](#), [pr#23324](#), Dan van der Ster)
- run-rbd-unit-tests.sh test fails to finish in jenkin’s make check run ([issue#27060](#), [issue#24910](#), [pr#23858](#), Mykola Golub)
- scrub livelock ([issue#26931](#), [issue#26890](#), [pr#23722](#), Sage Weil)
- spdk: compile with -march=core2 instead of -march=native ([issue#25032](#), [pr#23175](#), Nathan Cutler)
- tests: cluster [WRN] 25 slow requests in powercycle ([issue#25119](#), [pr#23886](#), Neha Ojha)
- test: Use pids instead of jobspecs which were wrong ([issue#32079](#), [issue#27056](#), [pr#23893](#), David Zafman)
- tools/ceph-detect-init: support RHEL as a platform ([issue#18163](#), [pr#23303](#), Nathan Cutler)
- tools: ceph-detect-init: support SLED ([issue#18163](#), [pr#23111](#), Nathan Cutler)
- tools: cephfs-data-scan: print the max used ino ([issue#26978](#), [issue#26925](#), [pr#23880](#), “Yan, Zheng”)

## v13.2.1 Mimic

---

This is the first bugfix release of the Mimic v13.2.x long term stable release series. This release contains many fixes across all components of Ceph, including a few security fixes. We recommend that all users upgrade.

## Notable Changes

---

- CVE 2018-1128: auth: cephx authorizer subject to replay attack ([issue#24836](#), Sage Weil)
- CVE 2018-1129: auth: cephx signature check is weak ([issue#24837](#), Sage Weil)
- CVE 2018-10861: mon: auth checks not correct for pool ops ([issue#24838](#), Jason Dillaman)

## Changelog

---

- bluestore: common/hobject: improved hash calculation for hobject\_t etc ([pr#22777](#), Adam Kupczyk, Sage Weil)
- bluestore,core: mimic: os/bluestore: don't store/use path\_block.{db,wal} from meta ([pr#22477](#), Sage Weil, Alfredo Deza)
- bluestore: os/bluestore: backport 24319 and 24550 ([issue#24550](#), [issue#24502](#), [issue#24319](#), [issue#24581](#), [pr#22649](#), Sage Weil)
- bluestore: os/bluestore: fix incomplete faulty range marking when doing compression ([pr#22910](#), Igor Fedotov)
- bluestore: spdk: fix ceph-osd crash when activate SPDK ([issue#24472](#), [issue#24371](#), [pr#22684](#), tone-zhang)
- build/ops: build/ops: ceph.git has two different versions of dpdk in the source tree ([issue#24942](#), [issue#24032](#), [pr#23070](#), Kefu Chai)
- build/ops: build/ops: install-deps.sh fails on newest openSUSE Leap ([issue#25065](#), [pr#23178](#), Kyr Shatskyy)
- build/ops: build/ops: Mimic build fails with -DWITH\_RADOSGW=0 ([issue#24766](#), [pr#22851](#), Dan Mick)
- build/ops: cmake: enable RTTI for both debug and release RocksDB builds ([pr#22299](#), Igor Fedotov)
- build/ops: deb/rpm: add python-six as build-time and run-time dependency ([issue#24885](#), [pr#22948](#), Nathan Cutler, Kefu Chai)
- build/ops: deb,rpm: fix block.db symlink ownership ([pr#23246](#), Sage Weil)
- build/ops: include: fix build with older clang (OSX target) ([pr#23049](#), Christopher Blum)
- build/ops: include: fix build with older clang ([pr#23034](#), Kefu Chai)
- build/ops, rbd: build/ops: order rbdmap.service before remote-fs-pre.target ([issue#24713](#), [issue#24734](#), [pr#22843](#), Ilya Dryomov)

- cephfs: cephfs: allow prohibiting user snapshots in CephFS ([issue#24705](#), [issue#24284](#), [pr#22812](#), "Yan, Zheng")
- cephfs: cephfs-journal-tool: Fix purging when importing an zero-length journal ([issue#24861](#), [pr#22981](#), yupeng chen, zhongyan gu)
- cephfs: client: fix bug #24491 \_ll\_drop\_pins may access invalid iterator ([issue#24534](#), [pr#22791](#), Liu Yangkuan)
- cephfs: client: update inode fields according to issued caps ([issue#24539](#), [issue#24269](#), [pr#22819](#), "Yan, Zheng")
- cephfs: common/DecayCounter: set last\_decay to current time when decoding dec... ([issue#24440](#), [issue#24537](#), [pr#22816](#), Zhi Zhang)
- cephfs,core: mon/MDSMonitor: do not send redundant MDS health messages to cluster log ([issue#24308](#), [issue#24330](#), [pr#22265](#), Sage Weil)
- cephfs: mds: add magic to header of open file table ([issue#24541](#), [issue#24240](#), [pr#22841](#), "Yan, Zheng")
- cephfs: mds: low wrlock efficiency due to dirfrags traversal ([issue#24704](#), [issue#24467](#), [pr#22884](#), Xuehan Xu)
- cephfs: PurgeQueue sometimes ignores Journaler errors ([issue#24533](#), [issue#24703](#), [pr#22810](#), John Spray)
- cephfs,rbd: osdc: Fix the wrong BufferHead offset ([issue#24583](#), [pr#22869](#), dongdong tao)
- cephfs: repeated eviction of idle client until some IO happens ([issue#24052](#), [issue#24296](#), [pr#22550](#), "Yan, Zheng")
- cephfs: test gets ENOSPC from bluestore block device ([issue#24238](#), [issue#24913](#), [issue#24899](#), [issue#24758](#), [pr#22835](#), Patrick Donnelly, Sage Weil)
- cephfs,tests: pjd: cd: too many arguments ([issue#24310](#), [pr#22882](#), Neha Ojha)
- cephfs,tests: qa: client socket inaccessible without sudo ([issue#24872](#), [issue#24904](#), [pr#23030](#), Patrick Donnelly)
- cephfs,tests: qa: fix ffsb cd argument ([issue#24719](#), [issue#24829](#), [issue#24680](#), [issue#24579](#), [pr#22956](#), Yan, Zheng, Patrick Donnelly)
- cephfs,tests: qa/suites: Add supported-random-distro\$ links ([issue#24706](#), [issue#24138](#), [pr#22700](#), Warren Usui)
- ceph-volume describe better the options for migrating away from ceph-disk ([pr#22514](#), Alfredo Deza)
- ceph-volume dmcrypt and activate -all documentation updates ([pr#22529](#), Alfredo

Deza)

- ceph-volume: error on commands that need ceph.conf to operate ([issue#23941](#), [pr#22747](#), Andrew Schoen)
- ceph-volume expand on the LVM API to create multiple LVs at different sizes ([pr#22508](#), Alfredo Deza)
- ceph-volume initial take on auto sub-command ([pr#22515](#), Alfredo Deza)
- ceph-volume lvm.activate Do not search for a MON configuration ([pr#22398](#), Wido den Hollander)
- ceph-volume lvm.common use destroy-new, doesn't need admin keyring ([issue#24585](#), [pr#22900](#), Alfredo Deza)
- ceph-volume: provide a nice error message when missing ceph.conf ([pr#22832](#), Andrew Schoen)
- ceph-volume tests destroy osds on monitor hosts ([pr#22507](#), Alfredo Deza)
- ceph-volume tests do not include admin keyring in OSD nodes ([pr#22425](#), Alfredo Deza)
- ceph-volume tests.functional install new ceph-ansible dependencies ([pr#22535](#), Alfredo Deza)
- ceph-volume: tests/functional run lvm list after OSD provisioning ([issue#24961](#), [pr#23148](#), Alfredo Deza)
- ceph-volume tests/functional use Ansible 2.6 ([pr#23244](#), Alfredo Deza)
- ceph-volume: unmount lvs correctly before zapping ([issue#24796](#), [pr#23127](#), Andrew Schoen)
- cmake: bump up the required boost version to 1.67 ([pr#22412](#), Kefu Chai)
- common: common: Abort in OSDMap::decode() during qa/standalone/erasure-code/test-erasure-eio.sh ([issue#24865](#), [issue#23492](#), [pr#23024](#), Sage Weil)
- common: common: fix typo in rados bench write JSON output ([issue#24292](#), [issue#24199](#), [pr#22406](#), Sandor Zeestraten)
- common,core: common: partially revert 95fc248 to make get\_process\_name work ([issue#24123](#), [issue#24215](#), [pr#22311](#), Mykola Golub)
- common: osd: Change osd\_skip\_data\_digest default to false and make it LEVEL\_DEV ([pr#23084](#), Sage Weil, David Zafman)
- common: tell ... config rm <foo> not idempotent ([issue#24468](#), [issue#24408](#), [pr#22552](#), Sage Weil)

- core: bluestore: flush\_commit is racy ([issue#24261](#), [issue#21480](#), [pr#22382](#), Sage Weil)
- core: ceph osd safe-to-destroy crashes the mgr ([issue#24708](#), [issue#23249](#), [pr#22805](#), Sage Weil)
- core: change default filestore\_merge\_threshold to -10 ([issue#24686](#), [issue#24747](#), [pr#22813](#), Douglas Fuller)
- core: common/hobject: improved hash calculation ([pr#22722](#), Adam Kupczyk)
- core: cosbench stuck at booting cosbench driver ([issue#24473](#), [pr#22887](#), Neha Ojha)
- core: librados: fix buffer overflow for aio\_exec python binding ([issue#24475](#), [pr#22707](#), Aleksei Gutikov)
- core: mon: enable level\_compaction\_dynamic\_level\_bytes for rocksdb ([issue#24375](#), [issue#24361](#), [pr#22361](#), Kefu Chai)
- core: mon/MgrMonitor: change ‘unresponsive’ message to info level ([issue#24246](#), [issue#24222](#), [pr#22333](#), Sage Weil)
- core: mon/OSDMonitor: no\_reply on MOSDFailure messages ([issue#24322](#), [issue#24350](#), [pr#22297](#), Sage Weil)
- core: os/bluestore: firstly delete db then delete bluefs if open db met error ([pr#22525](#), Jianpeng Ma)
- core: os/bluestore: fix races on SharedBlob::coll in ~SharedBlob ([issue#24859](#), [issue#24887](#), [pr#23065](#), Radoslaw Zarzynski)
- core: osd: choose\_acting loop ([issue#24383](#), [issue#24618](#), [pr#22889](#), Neha Ojha)
- core: osd: do not blindly roll forward to log.head ([issue#24597](#), [pr#22997](#), Sage Weil)
- core: osd: eternal stuck PG in ‘unfound\_recovery’ ([issue#24500](#), [issue#24373](#), [pr#22545](#), Sage Weil)
- core: osd: fix deep scrub with osd\_skip\_data\_digest=true (default) and blue... ([issue#24922](#), [issue#24958](#), [pr#23094](#), Sage Weil)
- core: osd: fix getting osd maps on initial osd startup ([pr#22651](#), Paul Emmerich)
- core: osd: increase default hard pg limit ([issue#24355](#), [pr#22621](#), Josh Durgin)
- core: osd: may get empty info at recovery ([issue#24771](#), [issue#24588](#), [pr#22861](#), Sage Weil)
- core: osd/PrimaryLogPG: rebuild attrs from clients ([issue#24768](#), [issue#24805](#), [pr#22960](#), Sage Weil)

- core: osd: retry to read object attrs at EC recovery ([issue#24406](#), [pr#22394](#), xiaofei cui)
- core: osd/Session: fix invalid iterator dereference in Sessoin::have\_backoff() ([issue#24486](#), [issue#24494](#), [pr#22730](#), Sage Weil)
- core: PG: add custom\_reaction Backfilled and release reservations after bac... ([issue#24332](#), [pr#22559](#), Neha Ojha)
- core: set correctly shard for existed Collection ([issue#24769](#), [issue#24761](#), [pr#22859](#), Jianpeng Ma)
- core,tests: Bring back diff -y for non-FreeBSD ([issue#24738](#), [issue#24470](#), [pr#22826](#), Sage Weil, David Zafman)
- core,tests: ceph\_test\_rados\_api\_misc: fix LibRadosMiscPool.PoolCreationRace ([issue#24204](#), [issue#24150](#), [pr#22291](#), Sage Weil)
- core,tests: qa/workunits/suites/blogbench.sh: use correct dir name ([pr#22775](#), Neha Ojha)
- core,tests: Wip scrub omap ([issue#24366](#), [issue#24381](#), [pr#22374](#), David Zafman)
- core,tools: ceph-detect-init: stop using platform.linux\_distribution ([issue#18163](#), [pr#21523](#), Nathan Cutler)
- core: ValueError: too many values to unpack due to lack of subdir ([issue#24617](#), [pr#22888](#), Neha Ojha)
- doc: ceph-bluestore-tool manpage not getting rendered correctly ([issue#25062](#), [issue#24800](#), [pr#23176](#), Nathan Cutler)
- doc: doc: update experimental features - snapshots ([pr#22803](#), Jos Collin)
- doc: fix the links in releases/schedule.rst ([pr#22372](#), Kefu Chai)
- doc: [mimic] doc/cephfs: remove lingering “experimental” note about multimds ([pr#22854](#), John Spray)
- lvm: when osd creation fails log the exception ([issue#24456](#), [pr#22640](#), Andrew Schoen)
- mgr/dashboard: Fix bug when creating S3 keys ([pr#22468](#), Volker Theile)
- mgr/dashboard: fix lint error caused by codelyzer update ([pr#22713](#), Tiago Melo)
- mgr/dashboard: Fix some datatable CSS issues ([pr#22274](#), Volker Theile)
- mgr/dashboard: Float numbers incorrectly formatted ([issue#24081](#), [issue#24707](#), [pr#22886](#), Stephan Müller, Tiago Melo)
- mgr/dashboard: Missing breadcrumb on monitor performance counters page

- ([issue#24764](#), [pr#22849](#), Ricardo Marques, Tiago Melo)
- mgr/dashboard: Replace Pool with Pools ([issue#24699](#), [pr#22807](#), Lenz Grimmer)
  - mgr: mgr/dashboard: Listen on port 8443 by default and not 8080 ([pr#22449](#), Wido den Hollander)
  - mgr,mon: exception for dashboard in config-key warning ([pr#22770](#), John Spray)
  - mgr,pybind: Python bindings use iteritems method which is not Python 3 compatible ([issue#24803](#), [issue#24779](#), [pr#22917](#), Nathan Cutler)
  - mgr: Sync up ceph-mgr prometheus related changes ([pr#22341](#), Boris Ranto)
  - mon: don't require CEPHX\_V2 from mons until nautilus ([pr#23233](#), Sage Weil)
  - mon/OSDMonitor: Respect paxos\_propose\_interval ([pr#22268](#), Xiaoxi CHEN)
  - osd: forward-port osd\_distrust\_data\_digest from luminous ([pr#23184](#), Sage Weil)
  - osd/OSDMap: fix CEPHX\_V2 osd requirement to nautilus, not mimic ([pr#23250](#), Sage Weil)
  - qa/rgw: disable testing on ec-cache pools ([issue#23965](#), [pr#23096](#), Casey Bodley)
  - qa/suites/upgrade/mimic-p2p: allow target version to apply ([pr#23262](#), Sage Weil)
  - qa/tests: added supported distro for powercycle suite ([pr#22224](#), Yuri Weinstein)
  - qa/tests: changed distro symlink to point to new way using supported OSes ([pr#22653](#), Yuri Weinstein)
  - rbd: librbd: deep\_copy: resize head object map if needed ([issue#24499](#), [issue#24399](#), [pr#22768](#), Mykola Golub)
  - rbd: librbd: fix crash when opening nonexistent snapshot ([issue#24637](#), [issue#24698](#), [pr#22943](#), Mykola Golub)
  - rbd: librbd: force 'invalid object map' flag on-disk update ([issue#24496](#), [issue#24434](#), [pr#22754](#), Mykola Golub)
  - rbd: librbd: utilize the journal disabled policy when removing images ([issue#24388](#), [issue#23512](#), [pr#22662](#), Jason Dillaman)
  - rbd: Prevent the use of internal feature bits from outside cls/rbd ([issue#24165](#), [issue#24203](#), [pr#22222](#), Jason Dillaman)
  - rbd: rbd-mirror daemon failed to stop on active/passive test case ([issue#24390](#), [pr#22667](#), Jason Dillaman)
  - rbd: [rbd-mirror] entries\_behind\_master will not be zero after mirror over ([issue#24391](#), [issue#23516](#), [pr#22549](#), Jason Dillaman)

- rbd: rbd-mirror simple image map policy doesn't always level-load instances ([issue#24519](#), [issue#24161](#), [pr#22892](#), Venky Shankar)
- rbd: rbd trash purge -threshold should support data pool ([issue#24476](#), [issue#22872](#), [pr#22891](#), Mahati Chamarthry)
- rbd,tests: qa: krbd\_exclusive\_option.sh: bump lock\_timeout to 60 seconds ([issue#25081](#), [pr#23209](#), Ilya Dryomov)
- rbd: yet another case when deep copying a clone may result in invalid object map ([issue#24596](#), [issue#24545](#), [pr#22894](#), Mykola Golub)
- rgw: cls\_bucket\_list fails causes cascading osd crashes ([issue#24631](#), [issue#24117](#), [pr#22927](#), Yehuda Sadeh)
- rgw: multisite: RGWSyncTraceNode released twice and crashed in reload ([issue#24432](#), [issue#24619](#), [pr#22926](#), Tianshan Qu)
- rgw: objects in cache never refresh after rgw\_cache\_expiry\_interval ([issue#24346](#), [issue#24385](#), [pr#22643](#), Casey Bodley)
- rgw: add configurable AWS-compat invalid range get behavior ([issue#24317](#), [issue#24352](#), [pr#22590](#), Matt Benjamin)
- rgw: Admin OPS Api overwrites email when user is modified ([issue#24253](#), [pr#22523](#), Volker Theile)
- rgw: fix gc may cause a large number of read traffic ([issue#24807](#), [issue#24767](#), [pr#22941](#), Xin Liao)
- rgw: have a configurable authentication order ([issue#23089](#), [issue#24547](#), [pr#22842](#), Abhishek Lekshmanan)
- rgw: index complete miss zones\_trace set ([issue#24701](#), [issue#24590](#), [pr#22818](#), Tianshan Qu)
- rgw: Invalid Access-Control-Request may bypass validate\_cors\_rule\_method ([issue#24809](#), [issue#24223](#), [pr#22935](#), Jeegn Chen)
- rgw: meta and data notify thread miss stop cr manager ([issue#24702](#), [issue#24589](#), [pr#22821](#), Tianshan Qu)
- rgw:-multisite: endless loop in RGWBucketShardIncrementalSyncCR ([issue#24700](#), [issue#24603](#), [pr#22815](#), cfanz)
- rgw: performance regression for luminous 12.2.4 ([issue#23379](#), [issue#24633](#), [pr#22929](#), Mark Kogan)
- rgw: radogw-admin reshards status command should print text for reshards ([issue#24834](#), [issue#23257](#), [pr#23021](#), Orit Wasserman)

- rgw: “radosgw-admin objects expire” always returns ok even if the pro... ([issue#24831](#), [issue#24592](#), [pr#23001](#), Zhang Shaowen)
- rgw: require -yes-i-really-mean-it to run radosgw-admin orphans find ([issue#24146](#), [issue#24843](#), [pr#22986](#), Matt Benjamin)
- rgw: REST admin metadata API paging failure bucket & bucket.instance: InvalidArgument ([issue#23099](#), [issue#24813](#), [pr#22933](#), Matt Benjamin)
- rgw: set cr state if aio\_read err return in RGWCloneMetaLogCoroutine:state\_send\_rest\_request ([issue#24566](#), [issue#24783](#), [pr#22880](#), Tianshan Qu)
- rgw: test/rgw: fix for bucket checkpoints ([issue#24212](#), [issue#24313](#), [pr#22466](#), Casey Bodley)
- rgw,tests: add unit test for cls bi list command ([issue#24736](#), [issue#24483](#), [pr#22845](#), Orit Wasserman)
- tests: mimic - qa/tests: Set ansible-version: 2.4 ([issue#24926](#), [pr#23122](#), Yuri Weinstein)
- tests: osd sends op\_reply out of order ([issue#25010](#), [pr#23136](#), Neha Ojha)
- tests: qa/tests - added overrides stanza to allow runs on ovh on rhel OS ([pr#23156](#), Yuri Weinstein)
- tests: qa/tests - added skeleton for mimic point to point upgrades testing ([pr#22697](#), Yuri Weinstein)
- tests: qa/tests: fix supported distro lists for ceph-deploy ([pr#23017](#), Vasu Kulkarni)
- tests: qa: wait longer for osd to flush pg stats ([issue#24321](#), [pr#22492](#), Kefu Chai)
- tests: tests: Health check failed: 1 MDSs report slow requests (MDS\_SLOW\_REQUEST) in powercycle ([issue#25034](#), [pr#23154](#), Neha Ojha)
- tests: tests: make test\_ceph\_argparse.py pass on py3-only systems ([issue#24825](#), [issue#24816](#), [pr#22988](#), Nathan Cutler)
- tests: upgrade/luminous-x: whitelist REQUEST\_SLOW for rados\_mon\_thrash ([issue#25056](#), [issue#25051](#), [pr#23164](#), Nathan Cutler)

# v13.2.0 Mimic

This is the first stable release of Mimic, the next long term release series.

## Major Changes from Luminous

- *Dashboard*:

- The (read-only) Ceph manager dashboard introduced in Ceph Luminous has been replaced with a new implementation inspired by and derived from the [openATTIC](#) Ceph management tool, providing a drop-in replacement offering a [number of additional management features](#).

- *RADOS*:

- Config options can now be centrally stored and managed by the monitor.
- The monitor daemon uses significantly less disk space when undergoing recovery or rebalancing operations.
- An *async recovery* feature reduces the tail latency of requests when the OSDs are recovering from a recent failure.
- OSD preemption of scrub by conflicting requests reduces tail latency.

- *RGW*:

- RGW can now replicate a zone (or a subset of buckets) to an external cloud storage service like S3.
- RGW now supports the S3 multi-factor authentication API on versioned buckets.
- The Beast frontend is no longer experimental, and is considered stable and ready for use.

- *CephFS*:

- Snapshots are now stable when combined with multiple MDS daemons.

- *RBD*:

- Image clones no longer require explicit *protect* and *unprotect* steps.
- Images can be deep-copied (including any clone linkage to a parent image and associated snapshots) to new pools or with altered data layouts.

- *Misc*:

- We have dropped the Debian builds for the Mimic release due to the lack of

GCC 8 in Stretch. We expect Debian builds to return with the release of Buster in early 2019, and hope to build a final Luminous release (and possibly later Mimic point releases) once Buster is available.

## Upgrading from Luminous

### Notes

- We recommend you avoid creating any RADOS pools while the upgrade is in process.
- You can monitor the progress of your upgrade at each stage with the `ceph versions` command, which will tell you what ceph version(s) are running for each type of daemon.

### Instructions

1. If your cluster was originally installed with a version prior to Luminous, ensure that it has completed at least one full scrub of all PGs while running Luminous. Failure to do so will cause your monitor daemons to refuse to join the quorum on start, leaving them non-functional.

If you are unsure whether or not your Luminous cluster has completed a full scrub of all PGs, you can check your cluster's state by running:

```
1. # ceph osd dump | grep ^flags
```

In order to be able to proceed to Mimic, your OSD map must include the `recovery_deletes` and `purged_snapdirs` flags.

If your OSD map does not contain both these flags, you can simply wait for approximately 24-48 hours, which in a standard cluster configuration should be ample time for all your placement groups to be scrubbed at least once, and then repeat the above process to recheck.

However, if you have just completed an upgrade to Luminous and want to proceed to Mimic in short order, you can force a scrub on all placement groups with a one-line shell command, like:

```
1. # ceph pg dump pgs_brief | cut -d " " -f 1 | xargs -n1 ceph pg scrub
```

You should take into consideration that this forced scrub may possibly have a negative impact on your Ceph clients' performance.

2. Make sure your cluster is stable and healthy (no down or recovering OSDs). (Optional, but recommended.)

3. Set the `noout` flag for the duration of the upgrade. (Optional, but recommended.):

```
1. # ceph osd set noout
```

4. Upgrade monitors by installing the new packages and restarting the monitor daemons.:

```
1. # systemctl restart ceph-mon.target
```

Once all monitors are up, verify that the monitor upgrade is complete by looking for the `mimic` feature string in the mon map. For example:

```
1. # ceph mon feature ls
```

should include `mimic` under persistent features:

```
1. on current monmap (epoch NNN)
2.   persistent: [kraken,luminous,mimic]
3.   required: [kraken,luminous,mimic]
```

5. Upgrade `ceph-mgr` daemons by installing the new packages and restarting with:

```
1. # systemctl restart ceph-mgr.target
```

Verify the `ceph-mgr` daemons are running by checking `ceph -s`:

```
1. # ceph -s
2.
3. ...
4. services:
5.   mon: 3 daemons, quorum foo,bar,baz
6.   mgr: foo(active), standbys: bar, baz
7. ...
```

6. Upgrade all OSDs by installing the new packages and restarting the `ceph-osd` daemons on all hosts:

```
1. # systemctl restart ceph-osd.target
```

You can monitor the progress of the OSD upgrades with the new `ceph versions` or `ceph osd versions` command:

```
1. # ceph osd versions
2. {
3.   "ceph version 12.2.5 (...) luminous (stable)": 12,
```

```

4.      "ceph version 13.2.0 (...) mimic (stable)": 22,
5. }

```

7. Upgrade all CephFS MDS daemons. For each CephFS file system,

- Reduce the number of ranks to 1. (Make note of the original number of MDS daemons first if you plan to restore it later.):

```

1. # ceph status
2. # ceph fs set <fs_name> max_mds 1

```

- Wait for the cluster to deactivate any non-zero ranks by periodically checking the status:

```
1. # ceph status
```

- Take all standby MDS daemons offline on the appropriate hosts with:

```
1. # systemctl stop ceph-mds@<daemon_name>
```

- Confirm that only one MDS is online and is rank 0 for your FS:

```
1. # ceph status
```

- Upgrade the last remaining MDS daemon by installing the new packages and restarting the daemon:

```
1. # systemctl restart ceph-mds.target
```

- Restart all standby MDS daemons that were taken offline:

```
1. # systemctl start ceph-mds.target
```

- Restore the original value of `max_mds` for the volume:

```
1. # ceph fs set <fs_name> max_mds <original_max_mds>
```

8. Upgrade all radosgw daemons by upgrading packages and restarting daemons on all hosts:

```
1. # systemctl restart radosgw.target
```

9. Complete the upgrade by disallowing pre-Mimic OSDs and enabling all new Mimic-only functionality:

```
1. # ceph osd require-osd-release mimic
```

10. If you set `noout` at the beginning, be sure to clear it with:

```
1. # ceph osd unset noout
```

11. Verify the cluster is healthy with `ceph health`.

# Upgrading from pre-Luminous releases (like Jewel)

---

You *must* first upgrade to Luminous (12.2.z) before attempting an upgrade to Mimic. In addition, your cluster must have completed at least one scrub of all PGs while running Luminous, setting the `recovery_deletes` and `purged_snapdirs` flags in the OSD map.

## Upgrade compatibility notes

---

These changes occurred between the Luminous and Mimic releases.

- *core:*
  - The `pg force-recovery` command will not work for erasure-coded PGs when a Luminous monitor is running along with a Mimic OSD. Please use the recommended upgrade order of monitors before OSDs to avoid this issue.
  - The sample `crush-location-hook` script has been removed. Its output is equivalent to the built-in default behavior, so it has been replaced with an example in the CRUSH documentation.
  - The `-f` option of the rados tool now means `--format` instead of `--force`, for consistency with the ceph tool.
  - The format of the `config diff` output via the admin socket has changed. It now reflects the source of each config option (e.g., default, config file, command line) as well as the final (active) value.
  - Commands variously marked as del, delete, remove etc. should now all be normalized as rm. Commands already supporting alternatives to rm remain backward-compatible. This changeset applies to the `radosgw-admin` tool as well.
  - Monitors will now prune on-disk full maps if the number of maps grows above a certain number (`mon_osdmap_full_prune_min`, default: 10000), thus preventing unbounded growth of the monitor data store. This feature is enabled by default, and can be disabled by setting `mon_osdmap_full_prune_enabled` to false.
  - *rados list-inconsistent-obj format changes:*
    - Various error strings have been improved. For example, the “oi” or “oi\_attr” in errors which stands for object info is now “info” (e.g. `oi_attr_missing` is now `info_missing`).
    - The object’s “selected\_object\_info” is now in json format instead of string.

- The attribute errors (attr\_value\_mismatch, attr\_name\_mismatch) only apply to user attributes. Only user attributes are output and have the internal leading underscore stripped.
  - If there are hash information errors (hinfo\_missing, hinfo\_corrupted, hinfo\_inconsistency) then “hashinfo” is added with the json format of the information. If the information is corrupt then “hashinfo” is a string containing the value.
  - If there are snapset errors (snapset\_missing, snapset\_corrupted, snapset\_inconsistency) then “snapset” is added with the json format of the information. If the information is corrupt then “snapset” is a string containing the value.
  - If there are object information errors (info\_missing, info\_corrupted, obj\_size\_info\_mismatch, object\_info\_inconsistency) then “object\_info” is added with the json format of the information instead of a string. If the information is corrupt then “object\_info” is a string containing the value.
- *rados list-inconsistent-snapset format changes:*
  - Various error strings have been improved. For example, the “ss\_attr” in errors which stands for snapset info is now “snapset” (e.g. ss\_attr\_missing is now snapset\_missing). The error snapset\_mismatch has been renamed to snapset\_error to better reflect what it means.
  - The head snapset information is output in json format as “snapset.” This means that even when there are no head errors, the head object will be output when any shard has an error. This head object is there to show the snapset that was used in determining errors.
- The osd\_mon\_report\_interval\_min option has been renamed to osd\_mon\_report\_interval, and the osd\_mon\_report\_interval\_max (unused) has been eliminated. If this value has been customized on your cluster then your configuration should be adjusted in order to avoid reverting to the default value.
  - The config-key interface can store arbitrary binary blobs but JSON can only express printable strings. If binary blobs are present, the ‘ceph config-key dump’ command will show them as something like <<< binary blob of length N >>>.
  - Bootstrap auth keys will now be generated automatically on a fresh deployment; these keys will also be generated, if missing, during upgrade.
  - The `osd force-create-pg` command now requires a force option to proceed because the command is dangerous: it declares that data loss is permanent and instructs the cluster to proceed with an empty PG in its place, without

making any further efforts to find the missing data.

### CephFS:

- Upgrading an MDS cluster to 12.2.3+ will result in all active MDS exiting due to feature incompatibilities once an upgraded MDS comes online (even as standby). Operators may ignore the error messages and continue upgrading/restarting or follow this upgrade sequence:

After upgrading the monitors to Mimic, reduce the number of ranks to 1 (`ceph fs set <fs_name> max_mds 1`), wait for all other MDS to deactivate, leaving the one active MDS, stop all standbys, upgrade the single active MDS, then upgrade/start standbys. Finally, restore the previous `max_mds`.

!! NOTE: see release notes on snapshots in CephFS if you have ever enabled snapshots on your file system.

See also: <https://tracker.ceph.com/issues/23172>

- Several `ceph mds ...` commands have been obsoleted and replaced by equivalent `ceph fs ...` commands:
  - `mds dump` -> `fs dump`
  - `mds getmap` -> `fs dump`
  - `mds stop` -> `mds deactivate`
  - `mds set_max_mds` -> `fs set max_mds`
  - `mds set` -> `fs set`
  - `mds cluster_down` -> `fs set cluster_down true`
  - `mds cluster_up` -> `fs set cluster_down false`
  - `mds add_data_pool` -> `fs add_data_pool`
  - `mds remove_data_pool` -> `fs rm_data_pool`
  - `mds rm_data_pool` -> `fs rm_data_pool`
- New CephFS file system attributes `session_timeout` and `session_autoclose` are configurable via `ceph fs set`. The MDS config options `mds_session_timeout`, `mds_session_autoclose`, and `mds_max_file_size` are now obsolete.
- As the multiple MDS feature is now standard, it is now enabled by default. `ceph fs set allow_multimds` is now deprecated and will be removed in a future release.
- As the directory fragmentation feature is now standard, it is now enabled by default. `ceph fs set allow_dirfrags` is now deprecated and will be removed in a

future release.

- MDS daemons now activate and deactivate based on the value of max\_mds. Accordingly, `ceph mds deactivate` has been deprecated as it is now redundant.
- Taking a CephFS cluster down is now done by setting the down flag which deactivates all MDS. For example: `ceph fs set cephfs down true`.
- Preventing standbys from joining as new actives (formerly the now deprecated cluster\_down flag) on a file system is now accomplished by setting the joinable flag. This is useful mostly for testing so that a file system may be quickly brought down and deleted.
- New CephFS file system attributes session\_timeout and session\_autoclose are configurable via `ceph fs set`. The MDS config options mds\_session\_timeout, mds\_session\_autoclose, and mds\_max\_file\_size are now obsolete.
- Each mds rank now maintains a table that tracks open files and their ancestor directories. Recovering MDS can quickly get open files' paths, significantly reducing the time of loading inodes for open files. MDS creates the table automatically if it does not exist.
- CephFS snapshot is now stable and enabled by default on new filesystems. To enable snapshot on existing filesystems, use the command:

```
1. ceph fs set <fs_name> allow_new_snaps
```

The on-disk format of snapshot metadata has changed. The old format metadata can not be properly handled in multiple active MDS configuration. To guarantee all snapshot metadata on existing filesystems get updated, perform the sequence of upgrading the MDS cluster strictly.

See <http://docs.ceph.com/docs/mimic/cephfs/upgrading/>

For filesystems that have ever enabled snapshots, the multiple-active MDS feature is disabled by the mimic monitor daemon. This will cause the “restore previous max\_mds” step in above URL to fail. To re-enable the feature, either delete all old snapshots or scrub the whole filesystem:

- `ceph daemon <mds of rank 0> scrub_path / force recursive repair`
- `ceph daemon <mds of rank 0> scrub_path '~mdsdir' force recursive repair`

- Support has been added in Mimic for quotas in the Linux kernel client as of v4.17.

See <http://docs.ceph.com/docs/mimic/cephfs/quota/>

- Many fixes have been made to the MDS metadata balancer which distributes load across MDS. It is expected that the automatic balancing should work well for most use-cases. In Luminous, subtree pinning was advised as a manual workaround for poor balancer behavior. This may no longer be necessary so it is recommended to try experimentally disabling pinning as a form of load balancing to see if the built-in balancer adequately works for you. Please report any poor behavior post-upgrade.
- NFS-Ganesha is an NFS userspace server that can export shares from multiple file systems, including CephFS. Support for this CephFS client has improved significantly in Mimic. In particular, delegations are now supported through the libcephfs library so that Ganesha may issue delegations to its NFS clients allowing for safe write buffering and coherent read caching. Documentation is also now available:  
<http://docs.ceph.com/docs/mimic/cephfs/nfs/>
- MDS uptime is now available in the output of the MDS admin socket `status` command.
- MDS performance counters for client requests now include average latency as well as the count.

- *RBD*

- The RBD C API’s `rbd_discard` method now enforces a maximum length of 2GB to match the C++ API’s `Image::discard` method. This restriction prevents overflow of the result code.
- The rbd CLI’s `lock list` JSON and XML output has changed.
- The rbd CLI’s `showmapped` JSON and XML output has changed.
- RBD now optionally supports simplified image clone semantics where non-protected snapshots can be cloned; and snapshots with linked clones can be removed and the space automatically reclaimed once all remaining linked clones are detached. This feature is enabled by default if the OSD “`require-min-compat-client`” flag is set to `mimic` or later; or can be overridden via the “`rbd_default_clone_format`” configuration option.
- RBD now supports deep copy of images that preserves snapshot history.

- *RGW*

- The RGW Beast frontend is now declared stable and ready for production use. [HTTP Frontends](#) for details.
- Civetweb frontend has been updated to the latest 1.10 release.
- The S3 API now has support for multi-factor authentication. Refer to [RGW Support for Multifactor Authentication](#) for details.

- RGW now has a sync plugin to sync to AWS and clouds with S3-like APIs.
- *MGR*
  - The (read-only) Ceph manager dashboard introduced in Ceph Luminous has been replaced with a new implementation, providing a drop-in replacement offering a number of additional management features. To access the new dashboard, you first need to define a username and password and create an SSL certificate. See the [Ceph Dashboard](#) for a feature overview and installation instructions.
  - The `ceph-rest-api` command-line tool (obsoleted by the MGR restful module and deprecated since v12.2.5) has been dropped.  
There is a MGR module called `restful` which provides similar functionality via a “pass through” method. See <http://docs.ceph.com/docs/master/mgr/restful> for details.
  - New command to track throughput and IOPS statistics, also available in `ceph -s` and previously in `ceph -w`. To use this command, enable the `iostat` Manager module and invoke it using `ceph iostat`. See the [iostat documentation](#) for details.
- *build/packaging*
  - The `rcceph` script (`systemd/ceph` in the source code tree, shipped as `/usr/sbin/rcceph` in the ceph-base package for CentOS and SUSE) has been dropped. This script was used to perform admin operations (start, stop, restart, etc.) on all OSD and/or MON daemons running on a given machine. This functionality is provided by the systemd target units (`ceph-osd.target`, `ceph-mon.target`, etc.).
  - The `python-ceph-compat` package is declared deprecated, and will be dropped when all supported distros have completed the move to Python 3. It has already been dropped from those supported distros where Python 3 is standard and Python 2 is optional (currently only SUSE).
  - Ceph codebase has now moved to the C++-17 standard.
  - The Ceph LZ4 compression plugin is now enabled by default, and introduces a new build dependency.

# Detailed Changelog

- arch/arm: set ceph\_arch\_aarch64\_crc32 only if the build host supports crc32cx ([issue#19705](#), [pr#17420](#), Kefu Chai)
- assert(false)->ceph\_abort() ([pr#18072](#), Li Wang)
- auth: keep /dev/urandom open for get\_random\_bytes ([issue#21401](#), [pr#17972](#), Casey Bodley)
- bluestore: BlueStore::ExtentMap::dup impl ([pr#19719](#), Shinobu Kinjo)
- bluestore: bluestore/NVMEDevice: accurate the latency perf counter of queue latency ([pr#17435](#), Ziye Yang, Pan Liu)
- bluestore: bluestore/NVMEDevice: convert the legacy config opt related with SPDK ([pr#18502](#), Ziye Yang)
- bluestore: bluestore/NVMEDevice: do not deference a dangling pointer ([pr#19067](#), Kefu Chai)
- bluestore: bluestore/NVMEDevice: fix the bug in write function ([pr#17086](#), Ziye Yang, Pan Liu)
- bluestore: bluestore/NVMeDevice: update NVMeDevice code due to SPDK upgrade ([pr#16927](#), Ziye Yang)
- bluestore,build/ops: bluestore,cmake: enable building bluestore without aio ([pr#19017](#), Kefu Chai)
- bluestore,build/ops: Build: create a proper WITH\_BLUESTORE option ([pr#18357](#), Alan Somers)
- bluestore,build/ops: ceph.spec.in,debian/rules: change aio-max-nr to 1048576 ([pr#17894](#), chenliuzhong)
- bluestore,build/ops,tests: os: add compile option to build libbluefs.so ([pr#16733](#), Pan Liu)
- bluestore,build/ops,tests: test/fio: fix build failure caused by sequencer replacement ([pr#20387](#), Igor Fedotov)
- bluestore: ceph-bluestore-tool: better fsck/repair, bluefs-bdev-{expand,sizes} ([pr#17709](#), Sage Weil)
- bluestore: ceph-bluestore-tool: check if bdev is empty on 'bluefs-bdev-expand' ([pr#17874](#), WANG Guoqin)
- bluestore: ceph-bluestore-tool: link target shouldn't ending with "n" ([pr#18585](#),

Yao Zongyou)

- bluestore,common: intarith: get rid of P2\* and ROUND\_UP\* macros ([pr#21085](#), xie xinggao)
- bluestore: comp\_min\_blob\_size init error ([pr#18318](#), linbing)
- bluestore: config: Change bluestore\_cache\_kv\_max to type INT64 ([pr#20255](#), Zhi Zhang)
- bluestore,core: ceph-bluestore-tool: prime-osd-dir: update symlinks instead of bailing ([pr#18565](#), Sage Weil)
- bluestore,core: common/options: bluefs\_buffered\_io=true by default ([pr#20542](#), Sage Weil)
- bluestore,core: os/bluestore: compensate for bad freelistmanager size/blocks metadata ([issue#21089](#), [pr#17268](#), Sage Weil)
- bluestore,core: os/bluestore: fix data read error injection in bluestore ([pr#19866](#), Sage Weil)
- bluestore,core: os/bluestore: kv\_max -> kv\_min ([pr#20544](#), Sage Weil)
- bluestore,core: os/bluestore: switch default allocator to stupid; test both bitmap and stupid in qa ([pr#16906](#), Sage Weil)
- bluestore,core: src/bluestore/NVMEDevice: make all read use aio\_submit ([pr#17655](#), Ziye Yang, Pan Liu)
- bluestore,core,tests: test/unittest\_bluefs: check whether rmdir success ([pr#15363](#), shiqi)
- bluestore,core: tool: ceph-kvstore-tool doesn't umount BlueStore properly ([issue#21625](#), [pr#18083](#), Chang Liu)
- bluestore: define default value of LoglevelV only once (3 templates) ([pr#20727](#), Matt Benjamin)
- bluestore: drop unused friend class in SharedDriverQueueData ([pr#16894](#), Pan Liu)
- bluestore: fix aio\_t::rval type ([issue#23527](#), [pr#21136](#), kungf)
- bluestore: fix build on armhf ([pr#20951](#), Kefu Chai)
- bluestore: fixed compilation error when enable spdk with gcc 4.8.5 ([pr#16945](#), Ziye Yang, Pan Liu)
- bluestore: kv/RocksDBStore: extract common code to a new function ([pr#16532](#), Pan Liu)
- bluestore/NVMEDevice: code cleanup ([pr#17284](#), Ziye Yang, Pan Liu)

- bluestore: os/bluestore: add bluestore\_prefer\_deferred\_size\_hdd/ssd to tracked keys ([pr#17459](#), xie xingguo)
- bluestore: os/bluestore: add discard method for ssd's performance ([pr#14727](#), Taeksang Kim)
- bluestore: os/bluestore: Add lat record of deferred\_queued and deferred\_aio\_wait ([pr#17015](#), lisali)
- bluestore: os/bluestore: Add missing \_\_func\_\_ in dout ([pr#17903](#), lisali)
- bluestore: os/bluestore: add perf counter for allocator fragmentation ([pr#21377](#), Igor Fedotov)
- bluestore: os/bluestore: allocate entire write in one go ([pr#17698](#), Sage Weil)
- bluestore: os/bluestore: allow reconstruction of osd data dir from bluestore bdev label ([pr#18256](#), Sage Weil)
- bluestore: os/bluestore: alter the allow\_eio policy regarding kernel's error list ([issue#23333](#), [pr#21306](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: avoid excessive ops in \_txc\_release\_alloc ([pr#18854](#), Igor Fedotov)
- bluestore: os/bluestore: avoid omit cache for remove-collection ([pr#18785](#), Jianpeng Ma)
- bluestore: os/bluestore: avoid overhead of std::function in blob\_t ([pr#20294](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: avoid unneeded BlobRefing in \_do\_read() ([pr#19864](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: be more verbose when hitting unloaded shard in extent map ([pr#21245](#), Igor Fedotov)
- bluestore: os/bluestore/BlueFS: compact log even when sync\_metadata sees no work ([pr#17354](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: Don't call debug related code under any condition ([pr#17627](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: don't need wait for aio when using \_sync\_write ([pr#16066](#), Haodong Tang)
- bluestore: os/bluestore/BlueFS: fix race with log flush during async log compaction ([issue#21878](#), [pr#18428](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: move release unused extents work in \_flush\_and\_syn\_log ([pr#17684](#), Jianpeng Ma)

- bluestore: os/bluestore/BlueFS: prevent \_compact\_log\_async reentry ([issue#21250](#), [pr#17503](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: Reduce unnecessary operations in collect\_metadata ([pr#17995](#), Luo Kexue)
- bluestore: os/bluestore/BlueFS: sanity check that alloc->allocate() won't return 0 ([pr#18259](#), xie xingguo)
- bluestore: os/bluestore/BlueFS: several cleanups ([pr#17966](#), xie xingguo)
- bluestore: os/bluestore/bluefs\_types: make block\_mask 64-bit ([pr#21629](#), Sage Weil)
- bluestore: os/bluestore/BlueStore: ASAP wake up \_kv\_finalize\_thread ([pr#18203](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueStore: narrow deferred\_lock in \_deferred\_submit\_unlock ([pr#17628](#), Jianpeng Ma)
- bluestore: os/bluestore: bluestore repair should use interval\_set::union\_insert ([pr#20900](#), Igor Fedotov)
- bluestore: os/bluestore: cleanup around ExtentList, AllocExtent and bluestore\_extent\_t classes ([pr#20360](#), Igor Fedotov)
- bluestore: os/bluestore: clearer comments, not slower code ([pr#16872](#), Mark Nelson)
- bluestore: os/bluestore: correctly check all block devices to decide if journal is\_rotational ([issue#23141](#), [pr#20602](#), Greg Farnum)
- bluestore: os/bluestore: delete redundant header file in KernelDevice.cc ([pr#18631](#), Jing Li)
- bluestore: os/bluestore: do not assert if BlueFS rebalance is unable to allocate sufficient space ([pr#18494](#), Igor Fedotov)
- bluestore: os/bluestore: do not core dump when BlueRocksEnv gets EEXIST error ([issue#20871](#), [pr#17357](#), liuchang0812)
- bluestore: os/bluestore: do not core dump when we try to open kvstore twice ([pr#18161](#), Chang Liu)
- bluestore: os/bluestore: do not release empty bluefs\_extents\_reclaiming ([pr#18671](#), Igor Fedotov)
- bluestore: os/bluestore: do not segv on kraken upgrade debug print ([issue#20977](#), [pr#16992](#), Sage Weil)
- bluestore: os/bluestore: don't re-initialize csum-setting for existing blobs

([issue#21175](#), [pr#17398](#), xie xingguo)

- bluestore: os/bluestore: do SSD discard on mkfs ([pr#20897](#), Igor Fedotov)
- bluestore: os/bluestore: drop deferred\_submit\_lock, fix aio leak ([issue#21171](#), [pr#17352](#), Sage Weil)
- bluestore: os/bluestore: drop unused function declaration ([pr#18075](#), Li Wang)
- bluestore: os/bluestore: drop unused param "what" in apply() ([pr#17251](#), songweibin)
- bluestore: os/bluestore: \_dump\_onode() don't prolongate Onode anymore ([pr#19841](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: dynamic CF configuration; put pglog omap in separate CF ([pr#18224](#), Sage Weil)
- bluestore: os/bluestore: enlarge aligned\_size avoid too many vector(> IOV\_MAX) ([issue#21932](#), [pr#18828](#), Jianpeng Ma)
- bluestore: os/bluestore: ExtentMap::reshard - fix wrong shard length ([pr#17334](#), chenliuzhong)
- bluestore: os/bluestore: fail early on very large objects ([issue#20923](#), [pr#16924](#), Sage Weil)
- bluestore: os/bluestore: fix another aio stall/deadlock ([issue#21470](#), [pr#18118](#), Sage Weil)
- bluestore: os/bluestore: fix broken cap in \_balance\_bluefs\_freespace() ([pr#21097](#), Igor Fedotov)
- bluestore: os/bluestore: fix clone dirty\_range again ([issue#20983](#), [pr#16994](#), Sage Weil)
- bluestore: os/bluestore: fix dirty\_shard off-by-one ([pr#16850](#), Sage Weil)
- bluestore: os/bluestore: fix exceeding the max IO queue depth in KernelDevice ([issue#23246](#), [pr#20996](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: fix potential assert when splitting collection ([pr#19519](#), Igor Fedotov)
- bluestore: os/bluestore: fix SharedBlob unregistration ([issue#22039](#), [pr#18805](#), Sage Weil)
- bluestore: os/bluestore: fix some code formatting ([pr#21037](#), Gu Zhongyan)
- bluestore: os/bluestore: fix the allocate in bluefs ([pr#19030](#), tangwenjun)
- bluestore: os/bluestore: fix the demotion in StupidAllocator::init\_rm\_free

([pr#20430](#), Kefu Chai)

- bluestore: os/bluestore: fix the wrong usage for map\_any ([pr#18939](#), Jianpeng Ma)
- bluestore: os/bluestore: fix wrong usage for BlueFS::\_allocate ([pr#20708](#), Jianpeng Ma)
- bluestore: os/bluestore: free the spdk qpair resource correctly in destructor of SharedDriverQueueData ([pr#20929](#), Jianyu Li)
- bluestore: os/bluestore: handle small main device properly ([pr#17416](#), xie xingguo)
- bluestore: os/bluestore: ignore 0x2000~2000 extent oddity from luminous upgrade ([issue#21408](#), [pr#17845](#), Sage Weil)
- bluestore: os/bluestore: implement BlueStore repair ([pr#19843](#), Igor Fedotov)
- bluestore: os/bluestore: make bluefs behave better near enospc ([pr#18120](#), Sage Weil)
- bluestore: os/bluestore: mark derivatives of AioContext as final ([pr#20227](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: move aio\_callback{,\_priv} to base class BlockDevice ([pr#17002](#), mychoxin)
- bluestore: os/bluestore: move assert of read/write to base class ([pr#17033](#), mychoxin)
- bluestore: os/bluestore: move size and block\_size to the base class BlockDevice ([pr#16886](#), Pan Liu)
- bluestore: os/bluestore: no need to fsync when failed to write label ([pr#20092](#), tangwenjun)
- bluestore: os/bluestore: no trim debug noise if there is no trimming to be done ([pr#20684](#), Sage Weil)
- bluestore: os/bluestore/NVMEDevice: change write\_bl to bl ([pr#17145](#), Ziye Yang, Pan Liu)
- bluestore: os/bluestore/NVMEDevice: fix the nvme queue depth issue ([pr#17200](#), Ziye Yang, Pan Liu)
- bluestore: os/bluestore/NVMEDevice: Remove using dpdk thread ([pr#17769](#), Ziye Yang, Pan Liu)
- bluestore: os/bluestore: OpSequencer: reduce kv\_submitted\_waiters if \_is\_all\_kv\_submitted() return true ([pr#18622](#), Jianpeng Ma)
- bluestore: os/bluestore: optimize \_collection\_list ([pr#18777](#), Jianpeng Ma)

- bluestore: os/bluestore: pass strict flag to bluestore\_blob\_use\_tracker\_t::equal() ([pr#15705](#), xie xingguo)
- bluestore: os/bluestore: Prealloc memory avoid realloc in list\_collection ([pr#18804](#), Jianpeng Ma)
- bluestore: os/bluestore: prevent mount if osd\_max\_object\_size >= 4G ([pr#19043](#), Sage Weil)
- bluestore: os/bluestore: print aio in batch ([pr#18873](#), Kefu Chai)
- bluestore: os/bluestore: print leaked extents to debug output ([pr#17225](#), Sage Weil)
- bluestore: os/bluestore: propagate read-EIO to high level callers ([pr#17744](#), xie xingguo)
- bluestore: os/bluestore: put cached attrs in correct mempool ([issue#21417](#), [pr#18001](#), Sage Weil)
- bluestore: os/bluestore: recalc\_allocated() when decoding bluefs\_fnode\_t ([issue#23212](#), [pr#20701](#), Jianpeng Ma, Kefu Chai)
- bluestore: os/bluestore: reduce meaningless flush ([pr#19027](#), tangwenjun)
- bluestore: os/bluestore: refactor FreeListManager to get clearer view on the number ([issue#22535](#), [pr#19718](#), Igor Fedotov)
- bluestore: os/bluestore: release disk extents in bulky manner ([pr#17913](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: remove ineffective BlueFS fnode extent calculation ([pr#18905](#), Igor Fedotov)
- bluestore: os/bluestore: remove unused parameters ([pr#18635](#), Jianpeng Ma)
- bluestore: os/bluestore: remove unused variable ([pr#21063](#), Gu Zhongyan)
- bluestore: os/bluestore: remove useless function submit ([pr#17537](#), mychoxin)
- bluestore: os/bluestore: reorder members of bluefs\_extent\_t for space efficiency ([pr#21034](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: replace dout with ldout in StupidAllocator ([pr#17404](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: report error and quit correctly when disk error happens ([issue#21263](#), [pr#17522](#), Pan Liu)
- bluestore: os/bluestore: Revert “os/bluestore: allow multiple DeferredBatches in flight at once” ([issue#20925](#), [issue#20295](#), [pr#16900](#), Sage Weil)

- bluestore: os/bluestore: s/bluefs\_total/bluefs\_free/ ([pr#21036](#), xie xingguo)
- bluestore: os/bluestore: separate finisher for deferred\_try\_submit ([issue#21207](#), [pr#17409](#), Sage Weil)
- bluestore: os/bluestore: set bitmap freelist resolution to min\_alloc\_size ([pr#17610](#), Sage Weil)
- bluestore: os/bluestore: shrink aio submit size to pending value ([pr#17588](#), kungf)
- bluestore: os/bluestore: silence -Wreturn-type warning ([pr#18286](#), Kefu Chai)
- bluestore: os/bluestore: support calculate cost when using spdk ([pr#17091](#), Ziye Yang, Pan Liu)
- bluestore: os/bluestore: synchronous on\_applied completions ([pr#18196](#), Sage Weil)
- bluestore: os/bluestore: trim cache every 50ms (instead of 200ms) ([pr#20498](#), Sage Weil)
- bluestore: os/bluestore: update description for bluestore\_compression\_[min|max|\_blob\_size options ([pr#21244](#), Igor Fedotov)
- bluestore: os/bluestore: using macro OBJECT\_MAX\_SIZE to check osd\_max\_object\_size ([pr#19622](#), Jianpeng Ma)
- bluestore: osd/bluestore: delete unused variable in KernelDevice ([pr#20857](#), Leo Zhang)
- bluestore: osd,os/bluestore: Display current size of osd\_max\_object\_size ([pr#19725](#), Shinobu Kinjo)
- bluestore: Revert “os/bluestore: pass strict flag to bluestore\_blob\_use\_tracker\_t::equal()” ([issue#21293](#), [pr#17569](#), Sage Weil)
- bluestore,rgw: rgw,unittest\_bit\_alloc: silence clang analyzer warning ([pr#17294](#), Kefu Chai)
- bluestore,tests: objectstore/store\_test: fix lack of flush prior to collection\_empty()... ([issue#22409](#), [pr#19764](#), Igor Fedotov)
- bluestore,tests: Revert “bluestore/fio: Fixed problem with all objects having the same hash ([pr#18352](#), Radoslaw Zarzynski)
- bluestore,tools: ceph-bluestore-tool: create out\_dir before create full path of kvdb ([pr#18367](#), Leo Zhang)
- bluestore,tools: os/bluestore/bluestore\_tool: add log-dump command to dump bluefs's log ([pr#18535](#), Yang Honggang)
- build: fix dpdk build error ([pr#18087](#), chunmei)

- build mimic-dev1 with gcc 7 ([issue#22438](#), [pr#19548](#), Kefu Chai)
- build/ops: automake: remove files required by automake ([pr#17937](#), Kefu Chai)
- build/ops: blkin: link against lttnq-ust-fork ([pr#17673](#), Mohamad Gebai)
- build/ops: boost: remove boost submodule ([pr#17405](#), Kefu Chai)
- build/ops: build: do\_cmake: allow ARGS to be overridden ([pr#19876](#), Abhishek Lekshmanan)
- build/ops: build: remove PGMap.cc from libcommon ([pr#18496](#), Sage Weil)
- build/ops: ceph-disk activate unlocks bluestore data partition ([issue#20488](#), [pr#16357](#), Felix Winterhalter)
- build/ops: ceph\_disk: allow “no fsid” on activate ([pr#18991](#), Dan Mick)
- build/ops,cephfs: ceph-object-corpus: update to fix make check ([pr#21261](#), Patrick Donnelly)
- build/ops,cephfs: cmake, test/fs, client: fix build with clang ([pr#20392](#), Adam C. Emerson)
- build/ops: ceph.spec: use devtoolset-6-gcc-c++ on aarch64 ([issue#22301](#), [pr#19341](#), Kefu Chai)
- build/ops: ceph-volume: Require lvm2, move to osd package ([issue#22443](#), [pr#19529](#), Theofilos Mouratidis)
- build/ops: ceph-volume: tests add tests for the is\_mounted utility ([pr#16962](#), Alfredo Deza)
- build/ops: change WITH\_SYSTEMD default to ON ([pr#20404](#), Nathan Cutler)
- build/ops: cmake/BuildBoost: fixes to ready seastar ([pr#20616](#), Kefu Chai, Casey Bodley)
- build/ops: cmake,deb: install system units using cmake ([pr#20618](#), Kefu Chai)
- build/ops: cmake: link libcommon with libstdc++ statically if WITH\_STATIC\_LIBSTDCXX ([issue#22438](#), [pr#19515](#), Kefu Chai)
- build/ops: cmake,make-dist: bump up boost version to 1.67 ([pr#21572](#), Kefu Chai)
- build/ops: cmake,mds: detect std::map::merge() before using it ([pr#21211](#), Willem Jan Withagen, Kefu Chai)
- build/ops: cmake/mgr: use Python 3 virtualenv if mgr subinterpreter is Python 3 ([pr#21446](#), Nathan Cutler)
- build/ops,common: cmake, common: silence cmake and gcc warnings ([issue#23774](#),

pr#21484, Kefu Chai)

- build/ops: common/time: add time.h for Alpine build (pr#19863, huanwen ren)
- build/ops,common: Update C++ standard to 14 and clean up (pr#19490, Adam C. Emerson)
- build/ops,core: ceph-crush-location: remove (pr#19881, Sage Weil)
- build/ops,core: ceph-volume: do not use -key during mkfs (issue#22283, pr#19276, Kefu Chai, Sage Weil)
- build/ops,core: /etc/sysconfig/ceph: remove jemalloc option (issue#20557, pr#18487, Sage Weil)
- build/ops,core: mimic: cmake,common,filestore: silence gcc-8 warnings/errors (pr#21862, Kefu Chai)
- build/ops,core: mimic: cmake: do not check for aligned\_alloc() anymore (issue#23653, pr#22048, Kefu Chai)
- build/ops,core: msg/async: update to work with dpdk shipped with spdk v17.10 (pr#19470, Kefu Chai)
- build/ops,core: zstd: Upgrade to v1.3.2 (pr#18407, Adam C. Emerson)
- build/ops: debian/control: adjust ceph-{osdomap,kvstore,monstore}-tool feature move (issue#22319, pr#19328, Sage Weil)
- build/ops: debian/control: adjust ceph-{osdomap,kvstore,monstore}-tool feature move (issue#22319, pr#19395, Kefu Chai, Sage Weil)
- build/ops: debian/control: adjust ceph-{osdomap,kvstore,monstore}-tool feature move (pr#19356, Kefu Chai)
- build/ops: debian: fix package relationships after 40caf6a6 (issue#21762, pr#18474, Kefu Chai)
- build/ops: debian: lock ceph user during purge (pr#15118, Caleb Boylan)
- build/ops: debian/rules: no more ChangeLog (pr#18023, Sage Weil)
- build/ops: debian/rules: strip ceph-base libraries (issue#22640, pr#19870, Sage Weil)
- build/ops: do\_{cmake,freebsd}: Don't invoke nproc(1) on FreeBSD (pr#17949, Alan Somers)
- build/ops: dpdk: remove redundant dpdk submodule (pr#18712, chunmei)
- build/ops: EventKqueue: Clang want realloc return to be typed (pr#21550, Willem Jan Withagen)

- build/ops: filestore,rgw: fix types/casts making clang on 32-Bit working ([pr#21055](#), Daniel Glaser)
- build/ops: Fix ppc64 support for ceph ([pr#16753](#), Boris Ranto)
- build/ops: Fix two dpdk assert happened in dpdk library ([pr#18409](#), chunmei)
- build/ops: FreeBSD: add new required packages to be installed ([pr#21349](#), Willem Jan Withagen)
- build/ops: githubmap: add some known Ceph reviewers ([pr#17507](#), Patrick Donnelly)
- build/ops: .githubmap: Add wjwithagen as a known Ceph reviewer ([pr#17518](#), Willem Jan Withagen)
- build/ops: .githubmap: Update ([pr#18230](#), Sage Weil)
- build/ops: .gitignore: allow debian .patch files ([pr#17577](#), Ken Dreyer)
- build/ops: include: compat.h, fix the return result of pthread\_set\_name() ([pr#20474](#), Willem Jan Withagen)
- build/ops: install-deps: Add support for ‘opensuse-tumbleweed’ ([pr#21650](#), Ricardo Marques)
- build/ops: install-deps.sh: avoid re-installing g++-7 ([pr#19468](#), Kefu Chai)
- build/ops: install-deps.sh, cmake: use GCC-7 on xenial also ([pr#19418](#), Kefu Chai)
- build/ops: install-deps.sh: install new gcc as the default the right way ([pr#19417](#), Kefu Chai)
- build/ops: install-deps.sh: pass -no-recommends to zypper ([issue#22998](#), [pr#20434](#), Nathan Cutler)
- build/ops: install-deps.sh: set python2 %bcond by environment ([issue#22999](#), [pr#20436](#), Nathan Cutler)
- build/ops: install-deps.sh: use DTS on centos if GCC is too old ([pr#19398](#), Kefu Chai)
- build/ops: install-deps.sh: use tee for writing a file ([pr#19516](#), Kefu Chai)
- build/ops: install-deps: use DTS-7 on aarch64 and only download mirrored package indexes ([pr#19645](#), Kefu Chai, Songbo Wang)
- build/ops: libmpem: Revert “submodule: make libmpem as a submodule.” ([pr#18414](#), Jianpeng Ma)
- build/ops: logrotate: add systemd reload in logrotate in case of centos minimal without killall ([pr#16586](#), Tianshan Qu)

- build/ops: make-dist,cmake: avoid re-downloading boost ([pr#19124](#), Kefu Chai)
- build/ops: make-dist,cmake: move boost tarball location to download.ceph.com ([pr#17980](#), Sage Weil)
- build/ops: make-dist,cmake: Try multiple URLs to download boost before failing ([pr#18048](#), Brad Hubbard)
- build/ops: make-dist: fall back to python3 ([pr#21127](#), Nathan Cutler)
- build/ops,mgr: mgr/dashboard: build tweaks ([pr#20752](#), John Spray)
- build/ops,mgr: mgr/dashboard: remove node/npm system installation ([pr#20898](#), Tiago Melo)
- build/ops,mgr: packaging: explicit jinja2 dependency for dashboard ([issue#22457](#), [pr#19598](#), John Spray)
- build/ops,mgr,tests: mgr/dashboard: replace dashboard with dashboard\_v2 ([pr#20912](#), Ricardo Dias)
- build/ops: mimic: cmake: use javac -h for creating JNI native headers ([issue#24012](#), [pr#21824](#), Kefu Chai)
- build/ops: mimic: silence various warnings to enable GCC-8 build ([pr#22081](#), Adam C. Emerson, Kefu Chai)
- build/ops: mon,osd: do not use crush\_device\_class file to initialize class for new osds ([pr#19939](#), Sage Weil)
- build/ops: mstart.sh: support read CLUSTERS\_LIST from env var ([pr#16988](#), Jiaying Ren)
- build/ops: os/CMakeLists: fix link erro when enable WITH\_PMEM=ON ([pr#20658](#), Jianpeng Ma)
- build/ops: osdc,os,osd: fix build on osx ([pr#20029](#), Kefu Chai)
- build/ops: python-numpy-devel build dependency for SUSE ([issue#21176](#), [pr#17366](#), Nathan Cutler)
- build/ops: qa/tests - added for the suites with subset be able to use 'testing' ... ([pr#21454](#), Yuri Weinstein)
- build/ops,rbd: ceph-dencoder: moved RBD types outside of RGW preprocessor guard ([issue#22321](#), [pr#19343](#), Jason Dillaman)
- build/ops: rbdmap: fix umount when multiple mounts use the same RBD ([pr#17978](#), Alexandre Marangone)
- build/ops: Revert "make-dist: add OBS-specific release suffix on SUSE" ([pr#20813](#), Nathan Cutler)

- build/ops, rgw: radosgw: Make compilation with CryptoPP possible ([pr#14955](#), Adam Kupczyk)
- build/ops: rocksdb: do not use aligned\_alloc ([issue#23653](#), [pr#21632](#), Kefu Chai)
- build/ops: rpm: adjust ceph-{osdomap,kvstore,monstore}-tool feature move ([issue#22558](#), [pr#19777](#), Kefu Chai)
- build/ops: rpm: build-depends on "cunit-devel" for suse ([pr#18997](#), Kefu Chai)
- build/ops: rpm: conditionalize Python 2 availability to enable Ceph build on Python 3-only system ([pr#20018](#), Nathan Cutler)
- build/ops: rpm,debian: Ensure all ceph-disk runtime dependencies are declared for ceph-base ([issue#23657](#), [pr#21356](#), Nathan Cutler)
- build/ops: rpm,deb: package ceph-kvstore-tool man page ([pr#17387](#), Sage Weil)
- build/ops: rpm: drop legacy librbd.so.1 symlink in /usr/lib64/qemu ([pr#17324](#), Nathan Cutler)
- build/ops: rpm: fix \_defined\_if\_python2\_absent conditional ([pr#20166](#), Nathan Cutler)
- build/ops: rpm: fix systemd macros for ceph-volume@.service ([issue#22217](#), [pr#19081](#), Nathan Cutler)
- build/ops: rpm: move ceph-\*-tool binaries out of ceph-test subpackage ([issue#21762](#), [pr#18289](#), Nathan Cutler)
- build/ops: rpm: Python 3-only ceph-disk and ceph-volume ([pr#20140](#), Nathan Cutler)
- build/ops: rpm: recommend chrony instead of ntp-daemon ([pr#20138](#), Nathan Cutler)
- build/ops: rpm: recommend python-influxdb with ceph-mgr ([pr#18511](#), Nathan Cutler, Tim Serong)
- build/ops: rpm: Revert "ceph.spec: work around build.opensuse.org" ([pr#21716](#), Nathan Cutler)
- build/ops: rpm: rip out rcceph script ([pr#19899](#), Nathan Cutler)
- build/ops: rpm: selinux-policy fixes ([pr#19026](#), Brad Hubbard)
- build/ops: rpm: set build parallelism based on available memory ([pr#19122](#), Nathan Cutler, Richard Brown)
- build/ops: rpm: set permissions 0755 on rbd resource agent ([issue#22362](#), [pr#19494](#), Nathan Cutler)
- build/ops: run-make-check.sh: fix SUSE support ([issue#22875](#), [pr#20234](#), Nathan Cutler)

- build/ops: run-make-check.sh: handle Python 2 absence ([issue#23035](#), [pr#20480](#), Nathan Cutler)
- build/ops: run-make-check.sh: run ulimit without sudo ([pr#17361](#), yang.wang)
- build/ops: script/build-integration-branch: print pr url list with titles ([pr#17426](#), Sage Weil)
- build/ops: selinux: Allow nvme devices ([issue#19200](#), [pr#15597](#), Boris Ranto)
- build/ops: setup-virtualenv.sh: do not hardcode python binary ([issue#23437](#), [pr#21002](#), Nathan Cutler)
- build/ops: spdk: update SPDK to fix the build failure on aarch64 ([pr#20134](#), Tone Zhang, Kefu Chai)
- build/ops: spdk: update SPDK to v17.10 ([pr#19208](#), Kefu Chai)
- build/ops: spdk: update submodule to more recent upstream ([pr#20077](#), Nathan Cutler)
- build/ops: specs: require of e2fsprogs ([pr#21345](#), Guillaume Abrioux)
- build/ops: src/script/build-integration-branch ([pr#17382](#), Sage Weil)
- build/ops: src: s/pip -use-wheel/pip/ ([pr#21159](#), Kefu Chai)
- build/ops: submodule: make libmpem as a submodule ([pr#17036](#), Jianpeng Ma)
- build/ops: sysctl.d: set kernel.pid\_max=4194304 ([issue#21929](#), [pr#18544](#), David Disseldorp)
- build/ops: systemd: rbd-mirror does not start on reboot ([pr#17969](#), Sébastien Han)
- build/ops: test: delete asok directories correctly ([pr#21023](#), Chang Liu)
- build/ops: test/fio: enable objectstore FIO plugin building without the need to install and build FIO source code ([pr#20535](#), Igor Fedotov)
- build/ops,tests: common,test,cmake: various changes to re-enable build on osx ([pr#18888](#), Kefu Chai)
- build/ops,tests: qa/tests: Changed rhel7.4 to rhel7.5 ([pr#21336](#), Yuri Weinstein)
- build/ops,tests: test/fio: fix fio objectstore plugin building broken by ([pr#20514](#), Igor Fedotov)
- build/ops: udev: Fix typo in udev OSD rules file ([pr#18976](#), Mitch Birti)
- build/ops: use devtoolset-7 on centos/rhel-7 ([pr#18863](#), Kefu Chai)
- cephfs: Client:Fix readdir bug ([pr#18784](#), dongdong tao)

- cephfs: Client: setattr should drop "Fs" rather than "As" for mtime and size ([pr#18786](#), dongdong tao)
- cephfs,common,rbd: common/common\_init: disable ms subsystem log gathering for clients ([issue#21860](#), [pr#18418](#), Jason Dillaman)
- cephfs,common,rbd: Various fixes for SCA issues ([pr#21708](#), Danny Al-Gaaf)
- cephfs,core: mon/OSDMonitor: set FLAG\_SELFMANAGED\_SNAPS on cephfs snap removal ([issue#23949](#), [pr#21756](#), Sage Weil)
- cephfs: MDS: add null check before we push\_back "onfinish" ([pr#18892](#), dongdong tao)
- cephfs: MDS: correct the error msg when init mon client ([pr#18836](#), dongdong tao)
- cephfs: MDS: make popular counter decay at proper rate ([pr#18776](#), Jianyu Li)
- cephfs: MDS: make rebalancer evaluate the overload state of each mds with the same criterion ([pr#19255](#), Jianyu Li)
- cephfs: messages: Initialization of is\_primary ([pr#16897](#), amitkuma)
- cephfs: messages: Initialization of member variables ([pr#16898](#), amitkuma)
- cephfs: mimic: MDSMonitor: clean up use of pending fsmap in uncommitted ops ([issue#23768](#), [pr#22005](#), Patrick Donnelly)
- cephfs: mon/MDSMonitor: wait for readable OSDMap before sanitizing ([issue#21945](#), [pr#18603](#), Patrick Donnelly)
- cephfs,mon: mon/MDSMonitor: fix a bug at preprocess\_beacon ([pr#17415](#), wangshuguang)
- cephfs: osdc/Journaler: use new style options ([pr#17806](#), Kefu Chai)
- cephfs: qa: check pool full flags ([issue#22475](#), [pr#19588](#), Patrick Donnelly)
- cephfs: qa: fix typo in test\_full ([issue#23643](#), [pr#21334](#), Patrick Donnelly)
- cephfs: Revert "ceph\_context: re-expand admin\_socket metavariables in child process" ([pr#18545](#), Patrick Donnelly)
- cephfs,tests: qa/suites/powercycle/osd/whitelist\_health: whitelist slow trimming ([pr#17307](#), Sage Weil)
- cephfs,tests: qa/workunits/cephtool/test.sh: fix test\_mon\_mds() ([pr#21579](#), Kefu Chai)
- cephfs,tools: mount.fuse.ceph: Fix typo ([pr#19128](#), Jos Collin)
- cephfs:vstart\_runner: fixes for recent cephfs changes ([pr#19533](#), Patrick

Donnelly)

- ceph-volume: add ANSIBLE\_SSH\_RETRIES=5 to functional tests ([pr#20592](#), Andrew Schoen)
- ceph-volume add functional tests for simple, rearrange lvm tests ([pr#18882](#), Alfredo Deza)
- ceph-volume: Add linesep/newline at end of JSON file when writing ([pr#19458](#), Wido den Hollander)
- ceph-volume: adds a --destroy flag to ceph-volume lvm zap ([issue#22653](#), [pr#20010](#), Andrew Schoen)
- ceph-volume: adds --crush-device-class flag for lvm prepare and create ([pr#19949](#), Andrew Schoen)
- ceph-volume: adds custom cluster name support to simple ([pr#20367](#), Andrew Schoen)
- ceph-volume: adds functional CI testing ([pr#16919](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: adds functional testing for bluestore ([pr#18656](#), Andrew Schoen)
- ceph-volume: adds raw device support to 'lvm list' ([issue#23140](#), [pr#20620](#), Andrew Schoen)
- ceph-volume: adds success messages for lvm prepare/activate/create ([issue#22307](#), [pr#19875](#), Andrew Schoen)
- ceph-volume: adds support to zap encrypted devices ([issue#22878](#), [pr#20537](#), Andrew Schoen)
- ceph-volume: adds the ceph-volume lvm zap subcommand ([pr#18513](#), Andrew Schoen)
- ceph-volume allow filtering by uuid, do not require osd id ([pr#17606](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: allow parallel creates ([issue#23757](#), [pr#21489](#), Theofilos Mouratidis)
- ceph-volume: allow skipping systemd interactions on activate/create ([issue#23678](#), [pr#21496](#), Alfredo Deza)
- ceph-volume: allow using a device or partition for lvm -data ([pr#18924](#), Alfredo Deza)
- ceph-volume be resilient to \$PATH issues ([pr#20650](#), Alfredo Deza)
- ceph-volume consume mount/format options from ceph.conf ([pr#20408](#), Alfredo Deza)
- ceph-volume: correctly fallback to bluestore when no objectstore is specified ([pr#19213](#), Alfredo Deza)

- ceph-volume correctly normalize mount flags ([pr#20543](#), Alfredo Deza)
- ceph-volume: create the ceph-volume and ceph-volume-systemd man pages ([issue#21030](#), [pr#17152](#), Alfredo Deza)
- ceph-volume: dmcrypt support for lvm ([issue#22619](#), [pr#20054](#), Alfredo Deza)
- ceph-volume dmcrypt support for simple ([issue#22620](#), [pr#20264](#), Andrew Schoen, Alfredo Deza)
- ceph-volume/doc: add missing subcommand in examples ([pr#19381](#), Guillaume Abrioux)
- ceph-volume: ensure correct -filestore/-bluestore behavior ([pr#18518](#), Alfredo Deza)
- ceph-volume failed ceph-osd -mkfs command doesn't halt the OSD creation process ([issue#23874](#), [pr#21685](#), Alfredo Deza)
- ceph-volume: fix action plugins path in tests ([pr#20910](#), Guillaume Abrioux)
- ceph-volume fix filestore OSD creation after mon-config changes ([issue#23260](#), [pr#20787](#), Alfredo Deza)
- ceph-volume: fix typo in ceph-volume lvm prepare help ([pr#21196](#), Jeffrey Zhang)
- ceph-volume: fix usage of the -osd-id flag ([issue#22642](#), [issue#22836](#), [pr#20203](#), Andrew Schoen)
- ceph-volume Format correctly when vg/lv cannot be used ([issue#22299](#), [pr#19285](#), Alfredo Deza)
- ceph-volume handle inline comments in the ceph.conf file ([issue#22297](#), [pr#19319](#), Alfredo Deza)
- ceph-volume: handle leading whitespace/tabs in ceph.conf ([issue#22280](#), [pr#19259](#), Alfredo Deza)
- ceph-volume Implement an 'activate all' to help with dense servers or migrating OSDs ([pr#21130](#), Alfredo Deza)
- ceph-volume improve robustness when reloading vms in tests ([pr#21070](#), Alfredo Deza)
- ceph-volume: log the current running command for easier debugging ([issue#23004](#), [pr#20594](#), Andrew Schoen)
- ceph-volume lvm api refactor/move ([pr#18110](#), Alfredo Deza)
- ceph-volume lvm list ([pr#18095](#), Alfredo Deza)
- ceph-volume lvm.prepare update to use create\_osd\_path ([pr#18514](#), Alfredo Deza)

- ceph-volume: lvm zap will unmount osd paths used by zapped devices ([issue#22876](#), [pr#20265](#), Andrew Schoen)
- ceph-volume: Nits noticed while studying code ([pr#21455](#), Dan Mick)
- ceph-volume Persist non-lv devices for journals ([pr#17403](#), Alfredo Deza)
- ceph-volume process the abspath of the executable first ([issue#23259](#), [pr#20824](#), Alfredo Deza)
- ceph-volume: removed the explicit use of sudo ([issue#22282](#), [pr#19363](#), Andrew Schoen)
- ceph-volume: remove extra space ([pr#21140](#), Sébastien Han)
- ceph-volume rollback on failed OSD prepare/create ([issue#22281](#), [pr#19351](#), Alfredo Deza)
- ceph-volume should be able to handle multiple LVM (VG/LV) tags ([issue#22305](#), [pr#19321](#), Alfredo Deza)
- ceph-volume: support GPT and other deployed OSDs ([pr#18823](#), Alfredo Deza)
- ceph-volume tests add optional flags for vagrant ([pr#20849](#), Alfredo Deza)
- ceph-volume tests alleviate libvirt timeouts when reloading ([issue#23163](#), [pr#20718](#), Alfredo Deza)
- ceph-volume tests.devices.lvm prepare isn't bluestore specific anymore ([pr#18984](#), Alfredo Deza)
- ceph-volume tests remove unused import ([pr#20459](#), Alfredo Deza)
- ceph-volume tests use granular env vars for vagrant ([pr#20864](#), Alfredo Deza)
- ceph-volume: Try to cast OSD metadata to int while scanning directory ([pr#19477](#), Wido den Hollander)
- ceph-volume update man page for prepare/activate flags ([pr#21570](#), Alfredo Deza)
- ceph-volume use realpath when checking mounts ([issue#22988](#), [pr#20427](#), Alfredo Deza)
- ceph-volume: use unique logical volumes ([pr#17207](#), Alfredo Deza)
- ceph-volume: Using -readonly for {vg|pv|lv}s commands ([pr#21409](#), Erwan Velu)
- ceph-volume: warn on missing ceph.conf file ([issue#22326](#), [pr#19347](#), Alfredo Deza)
- ceph-volume warn on mix of filestore and bluestore flags ([issue#23003](#), [pr#20513](#), Alfredo Deza)

- cleanup: Replacing MIN,MAX with std::min,std::max ([pr#18124](#), Amit Kumar)
- cli: rados: support for high precision time using stat2 ([issue#21199](#), [pr#17395](#), Abhishek Lekshmanan)
- cls\_acl/\_crypto: Add modeline ([pr#19010](#), Shinobu Kinjo)
- cmake: add chrono to BOOST\_COMPONENTS ([issue#23424](#), [pr#20977](#), Nathan Cutler)
- cmake: add cython\_rbd as a dependency to vstart target ([pr#18382](#), Ali Maredia)
- cmake: bail out if GCC version is less than 5.1 ([pr#19344](#), Kefu Chai)
- cmake: BuildBoost.cmake: use specified compiler for building boost ([pr#19898](#), Kefu Chai)
- cmake: bump target jdk to 1.7 ([issue#23458](#), [pr#21082](#), Shengjing Zhu)
- cmake: bump up required cmake version to 2.8.12 ([pr#18285](#), Kefu Chai)
- cmake: changes of BuildBoost.cmake to ready seastar ([pr#21404](#), Kefu Chai)
- cmake: check for aligned\_alloc() instead of checking tcmalloc version ([pr#18557](#), Kefu Chai)
- cmake: check gcc version not release date for libstdc++ saneness ([pr#18938](#), Kefu Chai)
- cmake: check version of boost in src/boost ([pr#19914](#), Kefu Chai)
- cmake: cleanups ([pr#18597](#), Kefu Chai)
- cmake,common: changes to port part of ceph to osx ([pr#17615](#), Kefu Chai)
- cmake: compile nvml as an external project ([pr#17462](#), Jianpeng Ma)
- cmake: define HAVE\_STDLIB\_MAP\_SPLICING for both libstdc++ and libc++ ([pr#21284](#), Kefu Chai)
- cmake: disable DOWNLOAD\_NO\_PROGRESS if cmake ver is lower than 3.1 ([pr#20492](#), Kefu Chai)
- cmake: disable FAIL\_ON\_WARNINGS for rocksdb ([pr#19426](#), Kefu Chai)
- cmake: disable VTA on options.cc ([pr#17393](#), Kefu Chai)
- cmake: do not find bzip2/lz4 for rocksdb ([pr#19963](#), runsisi)
- cmake: do not link against librados.a ([pr#18576](#), Kefu Chai)
- cmake: do not link against unused or duplicated libraries ([pr#18092](#), Kefu Chai)
- cmake: enabled py3 only build ([pr#20064](#), Kefu Chai)

- cmake: enable LZ4 by default ([pr#21332](#), Grant Slater, Casey Bodley)
- cmake: enable new policies to silence cmake warnings ([pr#21662](#), Kefu Chai)
- cmake: fix building without mgr module ([pr#21591](#), Yuan Zhou)
- cmake: fix frontend cmake build ([pr#21449](#), Ricardo Dias)
- cmake: fix libcephfs-test.jar build failure ([issue#22828](#), [pr#20175](#), Tone Zhang)
- cmake: fix the include dir for building boost::python ([pr#20324](#), Kefu Chai)
- cmake: fix typo in status message ([pr#21464](#), Lenz Grimmer)
- cmake: hide symbols import from other libraries in libcls\_\* ([issue#23517](#), [pr#21571](#), Kefu Chai)
- cmake: identify the possible incompatibility of rocksdb and tcmalloc ([issue#21422](#), [pr#17788](#), Kefu Chai)
- cmake: in case of bad "ALLOCATOR" selected issue warning ([pr#17422](#), Adam Kupczyk)
- cmake: include frontend build in 'all' target ([pr#21466](#), John Spray)
- cmake: let "tests" depend on "mgr-dashboard-frontend-build" ([pr#21468](#), Kefu Chai)
- cmake: 'make check' builds radosgw and its cls dependencies ([pr#20422](#), Casey Bodley)
- cmake: mgr: exclude .gitignore ([pr#19174](#), Nathan Cutler)
- cmake/modules/BuildRocksDB.cmake: enable compressions for rocksdb ([issue#24025](#), [pr#22183](#), Kefu Chai)
- cmake: only create sysctl file on linux ([pr#19029](#), Kefu Chai)
- cmake: pass static linkflags to the linker who links libcommon ([pr#19763](#), Kefu Chai)
- cmake: s/boost\_256/boost\_sha256/ ([pr#21573](#), Kefu Chai)
- cmake: set supported language the right way ([pr#18216](#), Kefu Chai)
- cmake: should use the value of GPERFTOOLS\_LIBRARIES as REQUIRED\_VARS ([pr#18645](#), Kefu Chai)
- cmake: s/sysconf/sysconfig/ ([pr#20631](#), Kefu Chai)
- cmake: sync nvml submodule to latest code ([pr#20411](#), Jianpeng Ma)
- cmake: System Includes to silence warnings from submodules and libraries! ([pr#18711](#), Adam C. Emerson)

- cmake: typo fix when npm is not found ([pr#20801](#), Abhishek Lekshmanan)
- cmake: update minimum boost version to 1.66 ([issue#20048](#), [issue#22600](#), [pr#19808](#), Casey Bodley)
- cmake: update the error message for gperftools bug ([pr#17901](#), Kefu Chai)
- cmake: warn if libstdc++ older than 5.1.0 is used ([pr#18837](#), Kefu Chai)
- cmake: WITH\_SPDK=ON by default ([pr#18944](#), Liu-Chunmei, Kefu Chai, wanjun.lp, Ziye Yang)
- common: adding line break at end of some cli results ([issue#21019](#), [pr#16687](#), songweibin)
- common: add line break for “ceph daemon TYPE.ID version” ([pr#17146](#), Zhu Shangzhong)
- common: Add metadata with only Ceph version number and release ([pr#21095](#), Wido den Hollander)
- common: Add min/max of ms\_async\_op\_threads ([pr#19942](#), Shinobu Kinjo)
- common: Add noreturn attribute to silence uninitialized warning ([pr#19348](#), Adam C. Emerson)
- common: auth: add err reason for log info in load function ([pr#17256](#), Luo Kexue)
- common: bench test fall into dead loop when <seconds>=0 ([pr#16382](#), PC)
- common: buffer: avoid changing bufferlist ABI by removing new \_mempool field ([issue#21573](#), [pr#18408](#), Sage Weil)
- common: by default, do not assert on leaks in the shared\_cache code ([issue#21737](#), [pr#18201](#), Greg Farnum)
- common: ceph: add the right bracket to watch-channel argument in the help message ([pr#19698](#), Chang Liu)
- common: ceph.in: execv using the same python ([pr#17713](#), Kefu Chai)
- common: ceph\_release: s/rc/stable/ ([pr#22264](#), Sage Weil)
- common: change routines to public access ([pr#20003](#), Willem Jan Withagen)
- common: Check this->data.op\_size before use ([pr#18816](#), Amit Kumar)
- common: cleanup address\_helper ([pr#19643](#), Shinobu Kinjo)
- common: cmake,common/RwLock: check for libpthread extensions ([pr#19202](#), Kefu Chai)

- common: common: add for\_each\_substr() for cheap string split ([pr#18798](#), Casey Bodley)
- common: common: add streaming interfaces for json/xml escaping ([pr#19806](#), Casey Bodley)
- common: common/admin\_socket: validate command json before feeding it to hook ([pr#20437](#), Kefu Chai)
- common: common/blkdev: fix build in FreeBSD environment ([pr#19316](#), Mykola Golub)
- common: common/buffer: cleanups ([pr#18312](#), Shinobu Kinjo)
- common: common/buffer: switch crc cache to single pair instead of map ([pr#18906](#), Piotr Dałek)
- common: common/config: add units to options ([issue#22747](#), [pr#20419](#), Kefu Chai)
- common: common/config: limit calls to normalize\_key\_name ([pr#20318](#), Piotr Dałek)
- common: common/config: make internal\_safe\_to\_start\_threads internal ([pr#18884](#), Sage Weil)
- common: common/ConfUtils: check key before actually normalizing ([pr#20370](#), Piotr Dałek)
- common: common/dns\_resolv.cc: Query for AAAA-record if ms\_bind\_ipv6 is True ([issue#23078](#), [pr#20530](#), Wido den Hollander)
- common: common/dns\_resolve: fix memory leak ([pr#19649](#), Yao Zongyou)
- common: common/event\_socket.h: include <errno.h> to use errno ([pr#18351](#), Kefu Chai)
- common: common/Formatter: fix string\_view usage for {json,xml}\_stream\_escaper ([issue#23622](#), [pr#21317](#), Sage Weil)
- common: common/hobject: compare two objects' key directly ([pr#21062](#), xie xingguo)
- common: common/hobject: preserve the order of hobject ([pr#21217](#), Kefu Chai)
- common: common/ipaddr: Do not select link-local IPv6 addresses ([issue#21813](#), [pr#20862](#), Wido den Hollander)
- common: common/lockdep: drop hash<pthread\_t> specialization ([pr#20574](#), Kefu Chai)
- common: common/LogClient: assign seq and queue atomically ([issue#18209](#), [pr#16828](#), Sage Weil)
- common: common/log: Speed improvement for log ([pr#19100](#), Adam Kupczyk, Kefu Chai)
- common: common/Ophistory: move insert/cleanup into separate thread ([pr#20540](#),

Piotr Dałek)

- common: common/options: drop unused options ([pr#20895](#), Kefu Chai)
- common: common/options: long description for log\_stderr\_prefix ([pr#19869](#), Sage Weil)
- common: common/options: pass by reference and use user-literals for size ([pr#18034](#), Kefu Chai)
- common: common/options: use user-defined literals for default values ([pr#17180](#), Kefu Chai)
- common: common/perf\_counters: remove unused parameter ([pr#19805](#), Kefu Chai)
- common: common/pick\_address.cc: Cleanup ([pr#19707](#), Shinobu Kinjo)
- common: common/pick\_address: wrong prefix\_len in pick\_iface() ([pr#20128](#), Gu Zhongyan)
- common: common/str\_list: s/boost::string\_view/std::string\_view ([pr#20475](#), Kefu Chai)
- common: common/strtol: fix strict\_strtoll() so it accepts hex starting with 0x ([pr#21521](#), Kefu Chai)
- common: common/strtoll: remove superfluous const modifier ([pr#21560](#), Jan Fajerski)
- common: common/throttle: start using 64-bit values ([issue#22539](#), [pr#19759](#), Igor Fedotov)
- common: common/types: make numbers a bit nicer when displaying space usage ([pr#17126](#), xie xingguo)
- common: common/util: do not print error if VERSION\_ID is missing ([pr#17787](#), Kefu Chai)
- common: compressor: use generate\_random\_number() for type="random" ([pr#18272](#), Casey Bodley)
- common: compressor/zstd: improvements ([pr#18879](#), Sage Weil)
- common: compute SimpleLRU's size with contents.size() instead of lru.size() ([issue#22613](#), [pr#19813](#), Xuehan Xu)
- common: config: expand tilde for ~/.ceph/\$cluster.conf ([issue#23215](#), [pr#20774](#), Rishabh Dave)
- common: config: notify config observers on set\_mon\_vals() ([pr#21161](#), Casey Bodley)

- common: config: Remove `_get_val` ([pr#18222](#), Adam C. Emerson)
- common/config: use `with_val()` for better performance ([pr#19056](#), Adam C. Emerson)
- common: consolidate spinlocks ([pr#15816](#), Jesse Williamson)
- common,core: common, osd: various cleanups ([pr#18149](#), Kefu Chai)
- common,core: common/pick\_address: add `{public,cluster}_network_interface` option ([pr#18028](#), Sage Weil)
- common,core: common/Throttle: Clean up ([pr#16618](#), Adam C. Emerson)
- common,core: fix broken use of `streamstream::rdbuf()` ([issue#22715](#), [pr#19998](#), Sage Weil)
- common,core: include/ceph\_features: deprecate a bunch of features ([pr#18546](#), Sage Weil)
- common,core: include,messages,rbd: Initialize counter,group\_pool ([pr#17774](#), Amit Kumar)
- common,core: options: Do not use linked lists of pointers! ([pr#17984](#), Adam C. Emerson)
- common,core: osdc/Objecter: take budgets across a LingerOp instead of on child Ops ([issue#22882](#), [pr#20519](#), Greg Farnum)
- common,core: osd/OpRequest: reduce overhead when disabling tracking ([pr#18470](#), Haomai Wang)
- common,core: rados: Prefer templates to macros ([pr#19913](#), Adam C. Emerson)
- common,core,rbd,rgw: common,osd,rgw: Fixes for issues found during SCA ([pr#21419](#), Danny Al-Gaaf)
- common,core,rbd,tests,tools: common,mds,mgr,mon,osd: store event only if it's added ([pr#16312](#), Kefu Chai)
- common,core: Revert "msg/async/AsyncConnection: unregister connection when racing happened" ([issue#22231](#), [pr#19586](#), Sage Weil)
- common,core: Revert "osd/OSDMap: allow bidirectional swap of pg-upmap-items" ([issue#21410](#), [pr#17760](#), Sage Weil)
- common: Coverity and SCA fixes ([pr#17431](#), Danny Al-Gaaf)
- common/crc/aarch64: Added cpu feature `pmull` and make aarch64 specific... ([pr#22184](#), Adam Kupczyk)
- common: crush/CrushWrapper: fix out of bounds access ([issue#20926](#), [pr#16869](#), Sage Weil)

- common: crypto: remove cryptopp library ([pr#20015](#), Casey Bodley)
- common: denc cleanups and other fixes ([pr#19877](#), Adam C. Emerson)
- common: denc: support enum with underlying type ([pr#18701](#), Kefu Chai)
- common: Destroy attr of RWLock after initialized ([pr#17103](#), Wen Zhang)
- common: dmclock: update mClockPriorityQueue with changes in subtree ([pr#20992](#), Casey Bodley)
- common: dout: DoutPrefixProvider operates directly on stream ([pr#21608](#), Casey Bodley)
- common: drop namespace using directives for std ([pr#19159](#), Shinobu Kinjo)
- common: drop unused variables "bluestore\_csum\_\*\_block" in opts ([pr#17394](#), songweibin)
- common: encoding: reset optional<> if it is uninitialized ([pr#17599](#), Kefu Chai)
- common: Extends random.h: numeric types relaxed to compatible types (with ([pr#20670](#), Jesse Williamson))
- common: fix BoundedKeyCounter const\_pointer\_iterator ([issue#22139](#), [pr#18953](#), Casey Bodley)
- common: fix daemon abnormal exit at parsing invalid arguments ([issue#21365](#), [issue#21338](#), [pr#17664](#), Yan Jun)
- common: fix potential memory leak in HTMLFormatter ([pr#20699](#), Yao Zongyou)
- common: fix typo deamon in comments ([pr#17687](#), yonghengdexin735)
- common: fix typo in options.cc ([pr#20549](#), songweibin)
- common: FreeBSD wants the correct struct selection for ipv6 ([issue#21813](#), [pr#21143](#), Willem Jan Withagen)
- common: global: output usage on -h, -help, or no args before contacting mons ([pr#20812](#), Sage Weil)
- common: hint the main branch of dout() accordingly to default verbosity ([pr#21259](#), Radoslaw Zarzynski)
- common: implement random number generator (following N3551) ([issue#18873](#), [pr#15341](#), Jesse Williamson)
- common: Improving message sent to user when getting signals ([issue#23320](#), [pr#21000](#), Erwan Velu)
- common: include/encoding: fix compat version error message ([pr#19660](#), Xinying

Song)

- common: include/interval\_set: parameterize by map type and kill btree\_interval\_set.h ([pr#18611](#), Sage Weil)
- common: include/rados: fix typo in librados.h ([pr#17988](#), wumingqiao)
- common: include: Remove unused header, ciso646 ([pr#18320](#), Shinobu Kinjo)
- common: include/types: format decimal numbers with decimal factor ([issue#22095](#), [pr#19117](#), Jan Fajerski)
- common: include: xlist: Fix Clang error for missing string ([pr#19367](#), Willem Jan Withagen)
- common: interval\_set: kill subset\_of() ([pr#21108](#), xie xingguo)
- common: interval\_set: optimize intersect\_of insert operations ([issue#21229](#), [pr#17265](#), Zac Medico)
- common: introduce md\_config\_cacher\_t ([pr#20320](#), Radoslaw Zarzynski)
- common: kick off mimic ([pr#16993](#), Sage Weil)
- common: lockdep fixes ([issue#20988](#), [pr#17738](#), Jeff Layton)
- common: log: clear thread-local stream's ios flags on reuse ([pr#20174](#), Casey Bodley)
- common: logically dead code inside shunique\_lock.h ([pr#17341](#), Amit Kumar)
- common: make ceph\_clock\_now() inlineable ([pr#20443](#), Radoslaw Zarzynski)
- common: Make code to invoke assert() smaller ([pr#20445](#), Adam Kupczyk)
- common: make some message informative, instead of error ([pr#16594](#), Willem Jan Withagen)
- common: mark events of TrackedOp outside its constructor ([issue#22608](#), [pr#19828](#), Xuehan Xu)
- common: mgr/dashboard\_v2: Fix test\_cluster\_configuration test ([issue#23265](#), [pr#20782](#), Sebastian Wagner)
- common: mimic: include/types: space between number and units ([pr#22107](#), Sage Weil)
- common,mon: crush,mon: fix weight-set vs crush device classes ([issue#20939](#), [pr#16883](#), Sage Weil)
- common,mon,osd,pybind: silence warning and remove executable mode bit ([pr#17512](#), Kefu Chai)

- common: msg/async/AsyncConnection: less noisy debug ([pr#20600](#), Sage Weil)
- common: msg/async: execute on core specified by core\_id not its index ([pr#20659](#), Kefu Chai)
- common: msg/msg\_types: fix the entity\_addr\_t's decoder ([pr#17699](#), Kefu Chai)
- common: msg/simple: s/ceph::size/std::size/ ([pr#19896](#), Kefu Chai)
- common/options.cc: cleanup readable literals for default sizes ([pr#18425](#), Enming Zhang)
- common/options.cc: Set Filestore rocksdb compaction readahead option ([issue#21505](#), [pr#17900](#), Mark Nelson)
- common: OpTracker doesn't visit TrackedOp when nref == 0 ([issue#24037](#), [pr#22160](#), Radoslaw Zarzynski)
- common: osdc/Objecter: fix warning ([pr#21757](#), Sage Weil)
- common: osdc/Objecter: record correctly value for l\_osdc\_op\_send\_bytes ([issue#21982](#), [pr#18810](#), Jianpeng Ma)
- common: osd/PrimaryLogPG: send requests to primary on cache miss ([issue#20919](#), [pr#16884](#), Sage Weil)
- common: osd\_types: define max in eversion\_t::max() to static ([pr#17453](#), yang.wang)
- common,os: initialize commit\_data, cmount, iocb ([pr#17766](#), Amit Kumar)
- common: posix\_fallocate on ZFS returns EINVAL ([pr#20398](#), Willem Jan Withagen)
- common: rados: clean up rados\_getxattrs() and rados\_striper\_getxattrs() ([pr#20259](#), Gu Zhongyan)
- common: RAII-styled mechanism for updating PerfCounters ([pr#19149](#), Radoslaw Zarzynski)
- common: random: revert change from boost::optional to std::optional ([issue#23778](#), [pr#21567](#), Casey Bodley)
- common: Remove ceph\_clock\_gettime, extern keyword ([pr#19353](#), Shinobu Kinjo)
- common: retry\_sys\_call no need take address of a function pointer ([pr#21281](#), Leo Zhang)
- common: Revert "common/config: return const reference instead of a copy" ([pr#18934](#), Kefu Chai)
- common: Revert "core: hint the dout()'s message crafting as a cold code." ([issue#23169](#), [pr#20636](#), Kefu Chai)

- common, rgw: rgw,common,rbd: s/boost::regex/std::regex/ ([pr#19393](#), Kefu Chai)
- common, rgw: rgw,common: remove already included header files ([pr#19390](#), Yao Zongyou)
- common: silence jenkins's buiding warning in obj\_bencher.cc ([pr#17272](#), Luo Kexue)
- common: src/common: update some ms\_\* options to be more consistent ([pr#20652](#), shangfufei)
- common: src/msg/async/rdma: decrease cpu usage by rdtsc instruction ([pr#16965](#), Jin Cai)
- common: Static Pointer ([pr#19079](#), Adam C. Emerson)
- common: strict\_strtol INT\_MAX and INT\_MIN is valid ([pr#18574](#), Shasha Lu)
- common: s/unique\_lock/lock\_guard/, if manual lock/unlock are not necessary ([pr#19770](#), Shinobu Kinjo)
- common: Switch singletons to use immobile\_any and cleanups ([pr#20273](#), Adam C. Emerson)
- common: test: fix unittest memory leak to silence valgrind ([pr#19654](#), Yao Zongyou)
- common, tests: test/common: unittest\_mclock\_priority\_queue builds with "make" command ([pr#17582](#), J. Eric Ivancich)
- common, tests: test/librados: create unique lock names ([issue#20798](#), [pr#16953](#), Neha Ojha)
- common: tools/crushtool: skip device id if no name exists ([issue#22117](#), [pr#18901](#), Jan Fajerski)
- common: use mono clock for HeartbeatMap ([pr#17827](#), Xinze Chi, Kefu Chai)
- common: use move instead of copy in build\_options() ([pr#18003](#), Casey Bodley)
- common: utime: fix \_\_32u sec time overflow ([pr#21113](#), kungf)
- compressor: add zstd back ([pr#21106](#), Kefu Chai)
- compressor: conditionalize on HAVE\_LZ4 ([pr#17059](#), Kefu Chai)
- compressor: kill AsyncCompressor which is broken ([pr#18472](#), Haomai Wang)
- core: blkin: Fix unconditional tracing in OSD ([pr#19156](#), Yingxin)
- core: ceph-debug-docker.sh: add ceph-osd-dbg package ([pr#17947](#), Patrick Donnelly)
- core: ceph.in: Add blocking mode for scrub and deep-scrub ([pr#19793](#), Brad

Hubbard)

- core: ceph.in: do not panic at control+d in interactive mode ([pr#18374](#), Kefu Chai)
- core: ceph.in: print all matched commands if arg missing ([issue#22344](#), [pr#19547](#), Kefu Chai)
- core: ceph.in: use a different variable for holding thrown exception ([pr#20663](#), Kefu Chai)
- core: ceph-kvstore-tool: copy to different store type and cleanup properly ([pr#18029](#), Josh Durgin)
- core: ceph-mgr: exit after usage ([issue#23482](#), [pr#21401](#), Sage Weil)
- core: ceph\_osd.cc: Drop legacy or redundant code ([pr#18718](#), Shinobu Kinjo)
- core: ceph-osd: some flags are not documented in the help output ([issue#20057](#), [pr#15565](#), Yanhu Cao)
- core: ceph: print output of “status” as string not as bytes ([pr#21297](#), Kefu Chai)
- core: ceph-rest-api: when port=0 use the DEFAULT\_PORT instead ([pr#17443](#), You Ji)
- core: ceph\_test\_objectstore: disable filestore\_fiemap for tests ([issue#21880](#), [pr#18452](#), Sage Weil)
- core: ceph\_test\_objectstore: do not change model for 0-length zero ([issue#21712](#), [pr#18519](#), Sage Weil)
- core: ceph\_test\_rados\_api\_aio: fix race with full pool and osdmap ([issue#23916](#), [issue#23917](#), [pr#21709](#), Sage Weil)
- core: ceph\_test\_rados\_api\_tier: add ListSnap test ([pr#17706](#), Xuehan Xu)
- core: client,osd,test: Initialize fuse\_req\_key,snap,who,seq ([pr#17772](#), Amit Kumar)
- core: common/admin\_socket: various cleanups ([pr#20274](#), Adam C. Emerson)
- core: common/config: cleanup remove some unused macros ([pr#19599](#), Yao Zongyou)
- core: common,mds,osd: Explicitly delete copy ctor if noncopyable ([pr#19465](#), Shinobu Kinjo)
- core: common/options: enable multiple rocksdb compaction threads for filestore ([pr#18232](#), Josh Durgin)
- core: common, osd: duplicated “start” event in OpTracker, improve OpTracker::dump\_ops ([pr#21119](#), Chang Liu)

- core: compressor: Add Brotli Compressor ([pr#19549](#), BI SHUN KE)
- core: config: lower default omap entries recovered at once ([issue#21897](#), [pr#19910](#), Josh Durgin)
- core: crush/CrushWrapper: fix potential invalid use of iterator ([pr#21325](#), xie xingguo)
- core: dmclock: Delivery of the dmclock delta, rho and phase parameter + Enabling the client service tracker ([pr#16369](#), Byungsu Park, Taewoong Kim)
- core: erasure-code: refactor the interfaces to hide internals from public ([pr#18683](#), Kefu Chai)
- core: erasure-code: use jerasure\_free\_schedule to properly free a schedule ([pr#19650](#), Yao Zongyou)
- core: erasure-code: use std::count() instead ([pr#19428](#), Kefu Chai)
- core: etc/default/ceph: remove jemalloc option ([issue#20557](#), [pr#18486](#), Sage Weil)
- core: filestore: include <linux/falloc.h> ([pr#20415](#), wumingqiao)
- core: Fix a dead lock when doing rdma performance test by fio ([pr#17016](#), Wang Chuanhong)
- core: Fix asserts caused by DNE pgs left behind after lots of OSD restarts ([issue#21833](#), [pr#20571](#), David Zafman)
- core: include: kill MIN and MAX macros ([pr#20886](#), Sage Weil)
- core: interval\_set: optimize intersection\_of ([pr#17088](#), Zac Medico)
- core: kv/KeyValueDB: add column family ([pr#18049](#), Jianjian Huo, Adam C. Emerson, Sage Weil)
- core: kv/RocksDB: get index and filter blocks memory usage ([pr#19934](#), Zhi Zhang)
- core: kv/RocksDBStore: fix rocksdb error when block cache is disabled ([issue#23816](#), [pr#21583](#), Yang Honggang)
- core: librados: add OPERATION\_ORDERNSNAP flag and yet another aio\_operate method ([pr#20343](#), Mykola Golub)
- core: librados.h: add LIBRADOS\_SUPPORTS\_APP\_METADATA ([pr#16542](#), Matt Benjamin)
- core: libradosstriper: fix the function declaration of rados\_striper\_trunc ([pr#20301](#), yuelongguang)
- core: libradosstriper: silence warning from -Wreorder ([pr#16890](#), songweibin)
- core: make the main dout() paths faster and more maintainable ([pr#20290](#), Radoslaw

Zarzynski)

- core: messages: Initialization of variable beat ([pr#17641](#), Amit Kumar)
- core: messages: Initialize member variables ([pr#16846](#), amitkuma)
- core: messages: initialize variable tid in MMDSFindIno ([pr#16793](#), amitkuma)
- core: messages: Initializing members in MOSDPGUpdateLogMissing ([pr#16928](#), amitkuma)
- core: messages: Initializing variable ceph\_mds\_reply\_head ([pr#17090](#), amitkuma)
- core: messages,journal: Initialization of stats\_period,m\_active\_set ([pr#17792](#), Amit Kumar)
- core: messages/MOSDMap: do compat reencode of crush map, too ([issue#21882](#), [pr#18454](#), Sage Weil)
- core: messages/MOSDOp: a fixes of encode\_payload ([pr#16836](#), Ying He)
- core: messages: Silence uninitialized member warnings ([pr#17596](#), Amit Kumar)
- core: mgr/DaemonServer.cc: add 'is\_valid=false' when decode caps error ([issue#20990](#), [pr#16978](#), Yanhu Cao)
- core,mgr: mgr/balancer: improve error message ([issue#22814](#), [pr#21427](#), Sage Weil)
- core,mgr: osd,mgrclient: pass daemon\_status by rvalue ref and other cleanups ([pr#18509](#), Kefu Chai)
- core,mgr: osd,mgr: report slow requests and pending creating pg to mgr ([pr#18614](#), Kefu Chai)
- core: mimic: crush: update choose\_args on bucket removal ([issue#24167](#), [pr#22120](#), Sage Weil)
- core: mimic: osdc: guard op->on\_notify\_finish with lock ([issue#23966](#), [pr#21834](#), Kefu Chai)
- core: mimic: osd: clean up smart probe ([issue#23899](#), [issue#24104](#), [pr#21959](#), Sage Weil, Gu Zhongyan)
- core: mimic: osd: Don't evict even when preemption has restarted with smaller chunk ([pr#22041](#), David Zafman)
- core: mimic: osd/PrimaryLogPG: fix try\_flush\_mark\_clean write contention case ([issue#24200](#), [issue#24174](#), [pr#22113](#), Sage Weil)
- core: mon/ConfigKeyService: dump: print placeholder value for binary blobs ([issue#23622](#), [pr#21329](#), Sage Weil)

- core,mon: crush, mon: bump up map version only if we truly created a weight-set ([pr#20178](#), xie xingguo)
- core: mon/LogMonitor: separate out summary by channel ([pr#21395](#), Sage Weil)
- core,mon: mon/AuthMonitor: create bootstrap keys on create\_initial() ([pr#21236](#), Joao Eduardo Luis)
- core,mon: mon/LogMonitor: do not crash on log sub w/ no messages ([pr#21469](#), Sage Weil)
- core,mon: mon,osd,crush: misc cleanup ([pr#20687](#), songweibin)
- core,mon: mon/OSDMonitor: Comment out unused function ([pr#20275](#), Brad Hubbard)
- core,mon: mon/OSDMonitor: don't create pgs if pool was deleted ([issue#21309](#), [pr#17600](#), Joao Eduardo Luis)
- core,mon: mon/OSDMonitor: implement cluster pg limit ([pr#17427](#), Sage Weil)
- core,mon: mon/OSDMonitor: list osd tree in named bucket ([pr#19564](#), kungf)
- core: mon, osd: add create-time for pool ([pr#21690](#), xie xingguo)
- core: mon, osd: fix potential collided \*Up Set\* after PG remapping ([issue#23118](#), [pr#20653](#), xie xingguo)
- core,mon: osd,mon: add max-pg-per-osd limit ([pr#18358](#), Kefu Chai)
- core: mon/OSDMonitor: filter out pgs that shouldn't exist from force-create-pg ([pr#20267](#), Sage Weil)
- core: mon/OSDMonitor: fix min\_size default for replicated pools ([pr#20555](#), Josh Durgin)
- core: mon/OSDMonitor: Fix OSDMonitor error message outputs ([issue#22351](#), [pr#20022](#), Brad Hubbard)
- core: mon/OSDMonitor: make 'osd crush class rename' idempotent ([pr#17330](#), xie xingguo)
- core: mon/OSDMonitor: rename outer name declaration to avoid shadowing ([pr#20032](#), Sage Weil)
- core: mon/OSDMonitor: require -yes-i-really-mean-it for force-create-pg ([pr#21619](#), Sage Weil)
- core: mon,osd,osdc: refactor snap trimming (phase 1) ([pr#18276](#), Sage Weil)
- core: mon, osd: per pool space-full flag support ([pr#17371](#), xie xingguo)
- core: mon, osd: turn down non-error scrub message severity ([issue#20947](#),

[pr#16916](#), John Spray)

- core: mon/PGMap: fix PGMapDigest decode ([pr#22099](#), Sage Weil)
- core: mon/PGMap: Fix %USED calculation bug ([issue#22247](#), [pr#19165](#), Xiaoxi Chen)
- core: mon/PGMap: remove or narrow columns 'pg ls' output ([pr#20945](#), Sage Weil)
- core: mon/PGMap: 'unclean' does not imply damaged ([pr#18493](#), Sage Weil)
- core: MOSDPGRecoveryDelete[Reply]: bump header version ([pr#17585](#), Josh Durgin)
- core: msg/asyc/rmda: fix the bug of assert when Infiniband::recv\_msg receives disconnect message ([pr#17688](#), Jin Cai)
- core: msg/async/AsyncConnection: combine multi alloc into one ([pr#18833](#), Haomai Wang)
- core: msg/async/AsyncConnection: Fix FPE in process\_connection ([issue#23618](#), [pr#21314](#), Brad Hubbard)
- core: msg/async/AsyncConnection: state will be NONE if replacing by another one ([issue#21883](#), [pr#18467](#), Haomai Wang)
- core: msg/async/AsyncConnection: unregister connection when racing happened ([pr#19013](#), Haomai Wang)
- core: msg/async: batch handle numevents ([pr#18321](#), Jianpeng Ma)
- core: msg/async: don't kill connection if replacing ([issue#21143](#), [pr#17288](#), Haomai Wang)
- core: msg/async: don't stuck into resetsession/retrysession loop ([pr#17276](#), Haomai Wang)
- core: msg/async: fix bug of data type conversion when uint64\_t -> int -> uint64\_t ([pr#18210](#), shangfufei)
- core: msg/async: print error log if add\_event fail ([pr#17102](#), mychoxin)

- core: msg/async/rdma: fix multi cephcontext confllicting ([pr#16893](#), Haomai Wang)
- core: msg/async/rdma: fix the bug that rdma polling thread uses the same thread name with msg worker ([pr#16936](#), Jin Cai)
- core: msg/async/rdma: improves RX buffer management ([pr#16693](#), Alex Mikheev)
- core: msg/async/rdma: uninitialized variable fix ([pr#18091](#), Vasily Philipov)
- core: msg/DispatchQueue: clear queue after wait() ([issue#18351](#), [pr#20374](#), Sage Weil)
- core: msgr/simple: set Pipe::out\_seq to in\_seq of the connecting side ([issue#23807](#), [pr#21585](#), Xuehan Xu)
- core: os/bluestore: debug bluestore cache shutdown ([issue#21259](#), [pr#17844](#), Sage Weil)
- core: os/bluestore: disable on\_applied sync\_complete ([issue#22668](#), [pr#20169](#), Sage Weil)
- core: os/bluestore: make bdev label parsing error more meaningful and less noisy ([pr#20090](#), Sage Weil)
- core: os/bluestore: make BlueStore opened by start\_kv\_only umountable ([issue#21624](#), [pr#18082](#), Chang Liu)
- core: os/bluestore: use db->rm\_range\_keys to delete range of keys ([pr#18279](#), Xiaoyan Li)
- core: OSD/admin\_socket: add get\_mapped\_pools command ([pr#19112](#), Xiaoxi Chen)
- core: osdc, class\_api: kill implicit string conversions ([pr#16648](#), Piotr Dałek)
- core: osdc: dec num\_in\_flight for pool\_dne case ([pr#21110](#), Jianpeng Ma)
- core: osdc: Do not use lock\_guard as unique\_lock ([pr#19756](#), Shinobu Kinjo)
- core: osdc: invoke notify finish context on linger commit failure ([issue#23966](#), [pr#21786](#), Jason Dillaman)
- core: osdc/Objecter: add ignore overlay flag if got redirect reply ([pr#21275](#), Ting Yi Lin)
- core: osdc/Objecter: delay initialization of hobject\_t in \_send\_op ([issue#21845](#), [pr#18427](#), Jason Dillaman)
- core: osdc/Objecter: fix recursive locking in \_finish\_command ([issue#23940](#), [pr#21742](#), Sage Weil)
- core: osdc/Objecter: misc cleanups ([pr#18476](#), Jianpeng Ma)

- core: osdc/Objecter: prevent double-invocation of linger op callback ([issue#23872](#), [pr#21649](#), Jason Dillaman)
- core: osdc/Objecter: skip sparse-read result decode if bufferlist is empty ([issue#21844](#), [pr#18400](#), Jason Dillaman)
- core: osd,compressor: Expose compression algorithms via MOSDBoot ([issue#22420](#), [pr#20558](#), Jesse Williamson)
- core: osdc: remove unused function ([pr#21081](#), Jianpeng Ma)
- core: osd,dmclock: use pointer to ClientInfo instead of a copy of it ([pr#18387](#), Kefu Chai)
- core: osd: do not forget pg\_stat acks which failed to send ([pr#16702](#), huangjun)
- core: OSD: drop unused parameter passed to check\_osdmap\_features ([pr#18466](#), Leo Zhang)
- core: osd/ECBackend: inject sleep during deep scrub ([pr#20531](#), xie xingguo)
- core: osd/ECBackend: only check required shards when finishing recovery reads ([issue#23195](#), [pr#21273](#), Josh Durgin)
- core: osd/ECBackend: update misleading comment about EIO handling ([pr#21686](#), Josh Durgin)
- core: osd/ECBackend: wait for apply for luminous peers ([pr#21604](#), Sage Weil)
- core: osd/ECMsgTypes: fix ECSubRead compat decode ([pr#20948](#), Sage Weil)
- core: osd, librados: add a rados op (TIER\_PROMOTE) ([pr#19362](#), Myoungwon Oh)
- core: osd,librados: add manifest, operations for chunked object ([pr#15482](#), Myoungwon Oh)
- core: osd,messages: Initialize read\_length,options,send\_reply ([pr#17799](#), Amit Kumar)
- core: osd/OSD: batch-list objects to reduce memory consumption ([pr#20767](#), xie xingguo)
- core: osd/OSD.cc: add 'isvalid=false' when failed to parse caps ([pr#16888](#), Yanhu Cao)
- core: osd/OSD.cc: use option 'osd\_scrub\_cost' instead ([pr#18479](#), Liao Weizhong)
- core: osd/OSDMap: add osdmap epoch info when printing info summary ([pr#20184](#), shun-s)
- core: osd/OSDMap: fix HAVE FEATURE logic in encode() ([pr#20922](#), Ilya Dryomov)

- core: osd/OSDMap: ignore PGs from pools of failure-domain OSD ([pr#20703](#), xie xingguo)
- core: osd/OSDMap: misleading message in print\_oneline\_summary() ([issue#22350](#), [pr#20313](#), Gu Zhongyan)
- core: osd/OSDMap: more pg upmap fixes ([issue#23878](#), [pr#21670](#), xiexingguo)
- core: osd/OSDMap: remove the unnecessary checks for null ([pr#18636](#), Kefu Chai)
- core: osd/OSDMap: skip out/crush-out osds ([pr#20655](#), xie xingguo)
- core: osd/OSDMap: upmap should respect the osd reweights ([issue#21538](#), [pr#17944](#), Theofilos Mouratidis)
- core: osd/osd\_type: get\_clone\_bytes - inline size() for overlapping size ([pr#17823](#), xie xingguo)
- core: osd/osd\_types.cc: copy extents map too while making clone ([pr#18396](#), xie xingguo)
- core: osd/osd\_types: fix ideal lower bound object-id of pg ([pr#21235](#), xie xingguo)
- core: osd/osd\_types: fix object\_stat\_sum\_t decode ([pr#18551](#), Sage Weil)
- core: osd/osd\_types: fix pg\_pool\_t encoding for hammer ([pr#21282](#), Sage Weil)
- core: osd/osd\_types: kill preferred field in pg\_t ([pr#20567](#), Sage Weil)
- core: osd/osd\_types: object\_info\_t: remove unused function ([pr#17905](#), Kefu Chai)
- core: osd/osd\_types: pg\_pool\_t: remove crash\_replay\_interval member ([pr#18379](#), Sage Weil)
- core: osd/osd\_types: remove backlog type for pg\_log\_entry\_t ([pr#20887](#), Sage Weil)
- core: osd/OSD: Using Wait rather than WaitInterval to wait queue.is\_empty() ([pr#17659](#), Jianpeng Ma)
- core: osd/PG: allow scrub preemption ([pr#18971](#), Sage Weil)
- core: osd/PGBBackend: delete reply if fails to complete delete request ([issue#20913](#), [pr#17183](#), Kefu Chai)
- core: osd/PGBBackend: drop input "snapid\_t" from objects\_list\_range() ([pr#21151](#), xie xingguo)
- core: osd/PGBBackend: fix large\_omap\_objects checking ([pr#21150](#), xie xingguo)
- core: osd/PGBBackend: release a msg using msg->put() not delete ([issue#20913](#), [pr#17246](#), Kefu Chai)

- core: osd/PG: const cleanup for recoverable/readable predicates ([pr#18982](#), Neha Ojha)
- core: osd/PG: decay scrub\_chunk\_max too if scrub is preempted ([pr#20552](#), xie xingguo)
- core: osd/PG: discard msgs from down peers ([issue#19605](#), [pr#17217](#), Kefu Chai)
- core: osd/PG: drop unused variable “oldest\_update” in PG.h ([pr#17142](#), songweibin)
- core: osd/PG: extend pg state bits to fix pg ls commands error ([issue#21609](#), [pr#18058](#), Yan Jun)
- core: osd/PG: fix calc of misplaced objects ([pr#18528](#), Kefu Chai)
- core: osd/PG: fix DeferRecovery vs AllReplicasRecovered race ([issue#23860](#), [pr#21706](#), Sage Weil)
- core: osd/PG: fix objects degraded higher than 100% ([issue#21803](#), [issue#21898](#), [pr#18297](#), Sage Weil, David Zafman)
- core: osd/PG: fix out of order priority for PG deletion ([pr#21613](#), xie xingguo)
- core: osd/PG: fix recovery op leak ([pr#18524](#), Sage Weil)
- core: osd/PG: fix uninit read in Incomplete::react(AdvMap&) ([issue#23980](#), [pr#21798](#), Sage Weil)
- core: osd/PG: force rebuild of missing set on jewel upgrade ([issue#20958](#), [pr#16950](#), Sage Weil)
- core: osd/PG: include primary in PG operator<< for ec pools ([pr#19453](#), Sage Weil)
- core: osd/PGLog: assert out on performing overflowed log trimming ([pr#21580](#), xie xingguo)
- core: osd/PGLog: cleanup unused function revise\_have ([pr#19329](#), Enming Zhang)
- core: osd/PGLog: fix sanity check against \*\*complete-to\*\* iter ([pr#21612](#), songweibin)
- core: osd/PGLog: get rid of ineffective container operations ([pr#19161](#), xie xingguo)
- core: osd/PGLog: write only changed dup entries ([issue#21026](#), [pr#17245](#), Josh Durgin)
- core: osd, pg, mgr: make snap trim queue problems visible ([issue#22448](#), [pr#19520](#), Piotr Dałek)
- core: osd/PG: misc cleanups ([pr#18340](#), Yan Jun)

- core: osd/PG: miscellaneous choose acting changes and cleanups ([pr#18481](#), xie xingguo)
- core: osd/PG: pass scrub priority to replica ([pr#20317](#), Sage Weil)
- core: osd/PG: prefer async\_recovery\_targets in reverse order of cost ([pr#21578](#), xie xingguo)
- core: osd/PG: prefer EC async\_recovery\_targets in reverse order of cost ([pr#21588](#), xie xingguo)
- core: osd/PG: PGPool::update: avoid expensive union\_of ([pr#17239](#), Zac Medico)
- core: osd/PGPool::update: optimize with subset\_of ([pr#17820](#), Zac Medico)
- core: osd/PG: reduce some overhead on operating MissingLoc ([pr#18186](#), xie xingguo)
- core: osd/PG: remote recovery preemption, and new feature bit to condition it on ([pr#18553](#), Sage Weil)
- core: osd/PG: remove unused parameter in calc\_ec\_acting ([pr#17304](#), yang.wang)
- core: osd/PG: restart recovery if NotRecovering and unfound found ([issue#22145](#), [pr#18974](#), Sage Weil)
- core: osd/PG: revert approx size ([issue#22654](#), [pr#18755](#), Adam Kupczyk)
- core: osd/PG: re-write of \_update\_calc\_stats and improve pg degraded state ([issue#20059](#), [pr#19850](#), David Zafman)
- core: osd/PG: some cleanups && add should\_gather filter for loop logging ([pr#19546](#), Enming Zhang)
- core: osd/PG: two cleanups ([pr#17171](#), xie xingguo)
- core: osd/PG: use osd\_backfill\_retry\_interval for schedule\_backfill\_retry() ([pr#18686](#), xie xingguo)
- core: osd/PrimaryLogPG: add condition "is\_chunky\_scrub\_active" to check object in chunky\_scrub ([pr#18506](#), Jianpeng Ma)
- core: osd/PrimaryLogPG: arrange recovery order by number of missing objects ([pr#18292](#), xie xingguo)
- core: osd/PrimaryLogPG: avoid infinite loop when flush collides with write lock ([pr#21653](#), Sage Weil)
- core: osd/PrimaryLogPG: calc clone\_overlap size in a more efficient and concise way ([pr#17928](#), xie xingguo)
- core: osd/PrimaryLogPG: cleanup do\_sub\_op && do\_sub\_op\_reply and define soiid in

- core: osd/PrimaryLogPG: prepare\_transaction more appropriate ([pr#19495](#), Enming Zhang)
- core: osd/PrimaryLogPG: clear data digest on WRITEFULL if skip\_data\_digest ([pr#21676](#), Sage Weil)
- core: osd/PrimaryLogPG: clear pin\_stats\_invalid bit properly on scrub-repair completion ([pr#18052](#), xie xingguo)
- core: osd/PrimaryLogPG: defer evict if head \*or\* object intersect scrub interval ([issue#23646](#), [pr#21628](#), Sage Weil)
- core: osd/PrimaryLogPG: do not pull-up snapc to snapset ([pr#18713](#), Sage Weil)
- core: osd/PrimaryLogPG: do not set data digest for bluestore ([pr#17515](#), xie xingguo)
- core: osd/PrimaryLogPG: do not set data/omap digest blindly ([pr#18061](#), xie xingguo)
- core: osd/PrimaryLogPG: do not use approx\_size() for log trimming ([pr#18338](#), xie xingguo)
- core: osd/PrimaryLogPG: do\_osd\_ops - propagate EAGAIN/EINPROGRESS on failok ([pr#17222](#), xie xingguo)
- core: osd/PrimaryLogPG: drop unused parameters ([pr#18581](#), Liao Weizhong)
- core: osd/PrimaryLogPG: fix dup stat for async read ([pr#18693](#), Xinze Chi)
- core: osd/PrimaryLogPG: Fix log messages ([pr#21639](#), Gu Zhongyan)
- core: osd/PrimaryLogPG: fix sparse read won't trigger repair correctly ([pr#17221](#), xie xingguo)
- core: osd/PrimaryLogPG: fix the oi size mismatch with real object size ([issue#23701](#), [pr#21408](#), Peng Xie)
- core: osd/PrimaryLogPG: kick off recovery on backoffing a degraded object ([pr#17987](#), xie xingguo)
- core: osd/PrimaryLogPG: kill add\_interval\_usage ([pr#17807](#), xie xingguo)
- core: osd/PrimaryLogPG: maybe\_handle\_manifest\_detail - sanity check obc existence ([pr#17298](#), xie xingguo)
- core: osd/PrimaryLogPG: misc cleanups ([pr#17830](#), Yan Jun)
- core: osd/PrimaryLogPG: more oi.extents fixes ([pr#18616](#), xie xingguo)
- core: osd/PrimaryLogPG: prepare\_transaction - fix EDQUOT vs ENOSPC ([pr#17808](#), xie xingguo)

- core: osd/PrimaryLogPG: request osdmap update in the right block ([issue#21428](#), [pr#17828](#), Josh Durgin)
- core: osd/PrimaryLogPG: several oi.extents fixes ([pr#18527](#), xie xingguo)
- core: osd/PrimaryLogPG: trigger auto-repair on full-object-size CRC error ([pr#18353](#), xie xingguo)
- core: osd/ReplicatedBackend: clean up code ([pr#20127](#), Jianpeng Ma)
- core: osd/ReplicatedBackend: ‘osd\_deep\_scrub\_keys’ doesn’t work ([pr#20221](#), fang yuxiang)
- core: osd/ReplicatedPG: add omap write bytes to delta\_stats ([pr#18471](#), Haomai Wang)
- core: osd\_types.cc: reorder fields in serialized pg\_stat\_t ([pr#19965](#), Piotr Dałek)
- core: os/filestore: disable rocksdb compression ([pr#18707](#), Sage Weil)
- core: os/filestore/FileStore: Initialized by nullptr, NULL or 0 instead ([pr#18980](#), Shinobu Kinjo)
- core: os/filestore: fix device/partition metadata detection ([issue#20944](#), [pr#16913](#), Sage Weil)
- core: os/filestore: fix do\_copy\_range replay bug ([issue#23298](#), [pr#20832](#), Sage Weil)
- core: os/Filestore: fix wbthrottle assert ([pr#14213](#), Xiaoxi Chen)
- core: os/filestore: print out the error if do\_read\_entry() fails ([pr#18346](#), Kefu Chai)
- core: os: FileStore, Using stl min | max, MIN | MAX macros instead ([pr#19832](#), Shinobu Kinjo)
- core: os: fix 0-length zero semantics, add tests ([issue#21712](#), [pr#18159](#), Sage Weil)
- core: os/FuseStore: fix incorrect used space statistics for fuse’s statfs interface ([pr#19033](#), Zhi Zhang)
- core: os/kstore: Drop unused function declaration ([pr#18077](#), Jos Collin)
- core: os/kstore: fix statfs problem and add vstart.sh support ([issue#23590](#), [pr#21287](#), Yang Honggang)
- core: os/memstore: Fix wrong use of lock\_guard ([pr#20914](#), Shen-Ta Hsieh)
- core: os/ObjectStore: fix get\_data\_alignment return -1 causing problem in msg

- header ([pr#18475](#), Haomai Wang)
- core: os/ObjectStore.h: fix mistake in comment TRANSACTION ISOLATION ([pr#16840](#), mychoxin)
  - core: os,osd: initial work to drop onreadable/onapplied callbacks ([issue#23029](#), [pr#20177](#), Sage Weil)
  - core: os: unify Sequencer and CollectionHandle ([pr#20173](#), Sage Weil)
  - core: PG: fix name of WaitActingChange ([pr#18768](#), wumingqiao)
  - core: pg: handle MNotifyRec event in down state ([pr#20959](#), Mingxin Liu)
  - core: PGPool::update: optimize removed\_snaps comparison when possible ([pr#17410](#), Zac Medico)
  - core: PGPool::update: optimize with interval\_set.swap ([pr#17121](#), Zac Medico)
  - core: PG: primary should not be in the peer\_info, skip if it is ([pr#20189](#), Neha Ojha)
  - core: ptl-tool: checkout branch after creation ([pr#18157](#), Patrick Donnelly)
  - core: ptl-tool: load labeled PRs ([pr#18231](#), Patrick Donnelly)
  - core: ptl-tool: make branch name configurable ([pr#18499](#), Patrick Donnelly)
  - core: ptl-tool: print bzs/tickets cited in commit msgs/comments ([pr#18547](#), Patrick Donnelly)
  - core: pybind/ceph\_argparse: fix cli output info ([pr#17667](#), Luo Kexue)
  - core: pybind/ceph\_argparse: Fix UnboundLocalError if command doesn't validate ([pr#21342](#), Tim Serong)
  - core: pybind/ceph\_argparse.py:'timeout' must in kwargs when call run\_in\_thread ([pr#21659](#), yangdeliu)
  - core,pybind: pybind/ceph\_argparse: accept flexible req ([pr#20791](#), Gu Zhongyan)
  - core,pybind: pybind/rados: add alignment getter to IoCtx ([pr#21222](#), Bruce Flynn)
  - core,pybind: pybind/rados: add rados\_service\_\*() ([pr#18812](#), Kefu Chai)
  - core,pybind: pybind/rados: add support open\_ioctx2 API ([pr#17710](#), songweibin)
  - core,pybind: rados: support python API of "set\_osdmap\_full\_try" ([pr#17418](#), songweibin)
  - core: qa: fix the potential delay of pg state change ([pr#17253](#), huangjun)
  - core: qa/standalone/osd/repro\_long\_log: no-mon-config for cot ([pr#20919](#), Sage

Weil)

- core: qa/standalone/scrub/osd-scrub-repair: no -y to diff ([issue#21618](#), [pr#18079](#), Sage Weil)
- core: qa/suite/rados: fix balancer vs firefly tunables failures ([pr#18826](#), Sage Weil)
- core: qa/suites/rados: fewer msgr failures ([pr#20918](#), Sage Weil)
- core: qa/suites/rados/perf: whitelist health warnings ([pr#18878](#), Sage Weil)
- core: qa/suites/rados/rest/mgr: provision openstack volumes ([pr#20573](#), Sage Weil)
- core: qa/suites/rados/singleton/all/mon-seesaw: whitelist MON\_DOWN ([pr#18246](#), Sage Weil)
- core: qa/suites/rados/singleton/all/recover-preemption: handle slow starting osd ([pr#18078](#), Sage Weil)
- core: qa/suites/rados/singleton/all/recovery\_preemption: whitelist SLOW\_OPS ([pr#21250](#), Sage Weil)
- core: qa/suites/rados/singleton/diverget\_priors\*: broaden whitelist ([pr#17379](#), Sage Weil)
- core: qa/suites/rados/thrash: extend mgr beacon grace when many msgr failures injected ([issue#21147](#), [pr#19242](#), Sage Weil)
- core: qa/suites/rados/verify/tasks/rados\_api\_tests: whitelist OBJECT\_MISPLACED ([pr#21646](#), Sage Weil)
- core: qa/workunits/rest/test.py: stop trying to test obsolete cluster\_up/down ([pr#18552](#), Sage Weil)
- core: rados/objclass.h: fix build define CEPH\_CLS\_API in all cases ([pr#21606](#), Danny Al-Gaaf)
- core: rados: use WaitInterval()'s return value instead of manual timing ([pr#20028](#), Mohamad Gebai)
- core, rbd: common, rbd-nbd: fix up prefork behavior vs AsyncMessenger singletons ([issue#23143](#), [pr#20681](#), Sage Weil)
- core, rbd: librbd, os: address coverity false positives ([pr#17793](#), Amit Kumar)
- core, rbd: mgr, osd, kv: Fix various warnings for Clang and GCC7 ([pr#17976](#), Adam C. Emerson)
- core, rbd: vstart.sh: fix mstart variables ([pr#20826](#), Sage Weil)
- core: rdma: Assign instead of compare ([pr#16664](#), amitkuma)

- core: remove startsync ([issue#20604](#), [pr#16396](#), Amit Kumar)
- core: rocksdb: sync with upstream ([issue#20529](#), [pr#17388](#), Kefu Chai)
- core: rocksdb: sync with upstream ([pr#21320](#), Kefu Chai)
- core: scrub errors not cleared on replicas can cause inconsistent pg state when replica takes over primary ([issue#23267](#), [pr#21101](#), David Zafman)
- core: Snapset inconsistency is detected with its own error ([issue#22996](#), [pr#20450](#), David Zafman)
- core: src/messages/MOSDMap: reencode OSDMap for older clients ([issue#21660](#), [pr#18134](#), Sage Weil)
- core: src/osd/PG.cc: 6455: FAILED assert(0 == "we got a bad state machine event") ([pr#20933](#), David Zafman)
- core: src/test/osd: add two pool test for manifest objects ([pr#20096](#), Myoungwon Oh)
- core: test/cli/osdmaptool/test-map-pgs.t: remove nondeterministic test ([pr#20872](#), Sage Weil)
- core: test/objectstore\_bench: Don't forget judging whether call usage ([pr#21369](#), Jianpeng Ma)
- core,tests: ceph\_test\_filestore\_idempotent\_sequence: many fixes ([issue#22920](#), [pr#20279](#), Sage Weil)
- core,tests: ceph\_test\_objectstore: drop expect regex ([pr#16968](#), Sage Weil)
- core,tests: Erasure code read test and code cleanup ([issue#14513](#), [pr#17703](#), David Zafman)
- core,tests: Erasure code recovery should send additional reads if necessary ([issue#21382](#), [pr#17920](#), David Zafman)
- core,tests: osd,dmclock: fix dmclock test simulator change ([pr#20270](#), J. Eric Ivancich)
- core,tests: os: kstore fix unittest for FiemapHole ([pr#17313](#), Ning Yao)
- core,tests: os/memstore: memstore\_page\_set=false ([issue#20738](#), [pr#16995](#), Sage Weil)
- core,tests: qa/ceph\_manager: check pg state again before timeout ([issue#21294](#), [pr#17810](#), huangjun)
- core,tests: qa/clusters/fixed-[23]: 4 osds per node, not 3 ([pr#16799](#), Sage Weil)
- core,tests: qa: modify rgw default pool names ([pr#21630](#), Neha Ojha)

- core, tests: qa/objectstore/bluestore\*: less debug output ([issue#20910](#), [pr#17505](#), Sage Weil)
- core, tests: qa/overrides/2-size-2-min-size: whitelist REQUEST\_STUCK ([pr#17243](#), Sage Weil)
- core, tests: qa/standalone/ceph-helpers: pass -verbose to ceph-disk ([pr#19456](#), Sage Weil)
- core, tests: qa/standalone/scrub/osd-scrub-repair: fix grep pattern ([issue#21127](#), [pr#17258](#), Sage Weil)
- core, tests: qa/standalone/scrub/osd-scrub-snaps: adjust test for lack of snapdir objects ([pr#17927](#), Sage Weil)
- core, tests: qa/suites/rados/monthrash: tolerate PG\_AVAILABILITY during mon thrashing ([pr#18122](#), Sage Weil)
- core, tests: qa/suites/rados/monthrash: whitelist SLOW\_OPS ([pr#21331](#), Sage Weil)
- core, tests: qa/suites/rados/objectstore: logs ([issue#20738](#), [pr#16923](#), Sage Weil)
- core, tests: qa/suites/rados/perf: create pool with lower pg\_num ([pr#17819](#), Neha Ojha)
- core, tests: qa/suites/rados/rest/mgr-restful: whitelist more health ([pr#18457](#), Sage Weil)
- core, tests: qa/suites/rados/rest: move rest\_test from qa/suites/rest/ ([pr#19175](#), Sage Weil)
- core, tests: qa/suites/rados/thrash: fix thrashing with ec vs map discon ([pr#16842](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: add hammer clients ([pr#21703](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: add rbd tests ([pr#21704](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: do some thrashing with jewel and luminous clients ([pr#21679](#), Sage Weil)
- core, tests: qa/suites/rados/thrash-old-clients: only centos and 16.04 ([pr#22125](#), Sage Weil)
- core, tests: qa/suites/upgrade/jewel-x/stress-split: tolerate sloppy past\_intervals ([pr#17226](#), Sage Weil)
- core, tests: qa/suites/upgrade/luminous-x/stress-split: avoid enospc ([pr#21753](#), Sage Weil)

- core, tests: qa/tasks/ceph\_manager: revive osds before doing final rerr reset ([issue#21206](#), [pr#17406](#), Sage Weil)
- core, tests: qa/tasks/ceph\_manager: tolerate tell osd.\* error ([pr#19365](#), Sage Weil)
- core, tests: qa/tasks/ceph.py: tolerate flush pg stats exception ([pr#16905](#), Sage Weil)
- core, tests: qa/tasks/filestore\_idempotent: shorter test ([pr#20509](#), Sage Weil)
- core, tests: qa/tasks/thrashosds: set min\_in default to 4 ([issue#21997](#), [pr#18670](#), Sage Weil)
- core, tests: qa/tests: run ceph-ansible task on installer.0 role/node ([pr#19605](#), Yuri Weinstein)
- core, tests: qa: tolerate failure to force backfill ([issue#22614](#), [pr#19765](#), Sage Weil)
- core, tests: qa/workunits/rados/test\_rados\_tool: fix stray  , race ([issue#22676](#), [pr#19946](#), Sage Weil)
- core, tests: qa/workunits/rados/test.sh: ensure tee output is valid filename ([pr#21507](#), Sage Weil)
- core, tests: rados: Initialization of alignment ([pr#17723](#), Amit Kumar)
- core, tests: rados: Initializing members of librados/TestCase.h ([pr#16896](#), amitkuma)
- core, tests: test: Checking fd for negative before closing ([pr#17190](#), amitkuma)
- core, tests: test: Check to avoid divide by zero ([pr#17220](#), amitkuma)
- core: tool: change default objectstore from filestore to bluestore ([pr#18066](#), Song Shun)
- core: tool: misc cleanup of ceph-kvstore-tool ([issue#22092](#), [pr#18815](#), Chang Liu)
- core, tools: Add export and remove ceph-objectstore-tool command option ([issue#21272](#), [pr#17538](#), David Zafman)
- core, tools: ceph-objectstore-tool: fix import of post-split pg from pre-split ancestor ([issue#21753](#), [pr#18229](#), Sage Weil)
- core: tools/ceph-objectstore-tool: split filestore directories offline to target hash level ([issue#21366](#), [pr#17666](#), Zhi Zhang)
- core, tools: common, tool: update kvstore-tool to repair key/value database ([issue#17730](#), [issue#21744](#), [pr#16745](#), liuchang0812, Chang Liu)

- core, tools: osd, os/bluestore: kill clang analyzer warnings ([pr#18015](#), Kefu Chai)
- core: tools/rados: add touch command to change object modification time ([pr#18913](#), Yao Zongyou)
- core, tools: scripts: add ptl-tool for scripting merges ([pr#17926](#), Patrick Donnelly)
- core: vstart.sh: drop .ceph\_port and use randomly selected available port ([pr#19268](#), Shinobu Kinjo)
- core: vstart.sh: drop -{mon,osd,mds,rgw,mgr}\_num options ([pr#18648](#), Kefu Chai)
- core: vstart.sh: Remove duplicate global section ([pr#17543](#), iliul)
- crush: cleanup update\_device\_class() log messages ([pr#21174](#), Gu Zhongyan)
- crush: fix CrushCompiler won't compile maps with empty shadow tree ([pr#17058](#), xie xingguo)
- crush: fix device\_class\_clone for unpopulated/empty weight-sets ([issue#23386](#), [pr#22169](#), Sage Weil)
- crush: fix fast rule lookup when uniform ([pr#17510](#), Sage Weil)
- crush: force rebuilding shadow hierarchy after swapping buckets ([pr#17083](#), xie xingguo)
- crush: improve straw2 algorithm's readability ([pr#20196](#), Yao Zongyou)
- crush: “osd crush class rename” support ([pr#16961](#), xie xingguo)
- crush, osd: handle multiple parents properly when applying pg upmaps ([issue#23921](#), [pr#21835](#), xiexingguo)
- crush: safe check for ‘ceph osd crush swap-bucket’ ([pr#17335](#), Carudy)
- crush: various CrushWrapper cleanups ([pr#17360](#), Kefu Chai)
- crush: various weight-set fixes ([pr#17014](#), xie xingguo)
- denc: should check element's type not ‘size\_t’ ([pr#19986](#), Kefu Chai)
- denc: use constexpr-if to replace some SFINAE impls ([pr#19662](#), Kefu Chai)
- doc: 12.1.3 release notes ([pr#16975](#), Abhishek Lekshmanan)
- doc: 12.2.0 major release announcements ([pr#16915](#), Abhishek Lekshmanan)
- doc: 12.2.1 release notes ([pr#18014](#), Abhishek Lekshmanan)
- doc: 12.2.4 release notes ([pr#20619](#), Abhishek Lekshmanan)

- doc: add 12.2.2 release notes ([pr#19264](#), Abhishek Lekshmanan)
- doc: add allow\_multimds and fs\_name parameter ([pr#15847](#), Jan Fajerski)
- doc: add ceph-kvstore-tool's man ([pr#17092](#), liuchang0812)
- doc: add changelog for 12.2.1 ([pr#18020](#), Abhishek Lekshmanan)
- doc: add changelog for v11.2.1 ([pr#16956](#), Abhishek Lekshmanan)
- doc: add changelog for v12.2.2 ([pr#19284](#), Abhishek Lekshmanan)
- doc: Added CHAP configuration instructions for iSCSI ([pr#18423](#), Ashish Singh)
- doc: add example of setting pool in cephfs layout ([pr#17372](#), John Spray)
- doc: Adding changelog for 10.2.10 ([pr#18151](#), Abhishek Lekshmanan)
- doc: Add introduction about different way to run rbd-mirror ([pr#19692](#), Yu Shengzuo)
- doc: add -max-buckets to radosgw-admin(8) ([pr#17439](#), Clément Pellegrini)
- doc: add missing blank line ([pr#18724](#), iliul)
- doc: Add missing pg states from doc ([pr#20504](#), David Zafman)
- doc: add mount.fuse.ceph to index ([issue#22595](#), [pr#19792](#), Jos Collin)
- doc: Add newbie-friendly updates to Helm start doc ([pr#18886](#), Blaine Gardner)
- doc: add osd\_max\_object\_size in osd configuration ([pr#18115](#), Mohamad Gebai)
- doc: build-doc: Upgrade ceph python libraries ([pr#20726](#), Boris Ranto)
- doc: ceph-disk: create deprecation warnings ([issue#22154](#), [pr#18988](#), Alfredo Deza)
- doc: ceph-volume: automatic VDO detection ([issue#23581](#), [pr#21451](#), Alfredo Deza)
- doc: ceph-volume docs ([pr#17068](#), Alfredo Deza)
- doc: ceph-volume document multipath support ([pr#20878](#), Alfredo Deza)
- doc: ceph-volume doc updates ([pr#20758](#), Alfredo Deza)
- doc: ceph-volume include physical devices associated with an LV when listing ([pr#21645](#), Alfredo Deza)
- doc: ceph-volume lvm bluestore support ([pr#18448](#), Alfredo Deza)
- doc/ceph-volume OSD use the fsid file, not the osd\_fsid ([issue#22427](#), [pr#20059](#), Alfredo Deza)
- doc: change boolean option default value from zero to false ([pr#17733](#), Yao

Zongyou)

- doc: change cn mirror to ustc domain ([pr#18081](#), Shengjing Zhu)
- doc: changelog for v12.2.3 ([pr#20503](#), Abhishek Lekshmanan)
- doc: cleanup erasure coded pool doc on cephfs use ([pr#20572](#), Patrick Donnelly)
- doc: CodingStyle: add python and javascript/typescript ([pr#20186](#), Joao Eduardo Luis)
- doc: common/options: document filestore and filejournal options ([pr#17739](#), Sage Weil)
- doc: common/options: document objecter, filer, and journal options ([pr#17740](#), Sage Weil)
- doc: complete and update the subsystem logging level info table ([pr#18500](#), Luo Kexue)
- doc: correcting typos in bluestore-config-ref and bluestore-migration ([pr#19154](#), Katie Holly)
- doc: correct wrong bluestore config types ([pr#18205](#), Yao Zongyou)
- doc: delete duplicate words ([pr#17104](#), iliul)
- doc: dev description of async recovery ([pr#21051](#), Neha Ojha, Josh Durgin)
- doc: doc/bluestore: add SPDK usage for bluestore ([pr#17654](#), Haomai Wang)
- doc: doc/cephfs/experimental-features: kernel client snapshots limit ([pr#18579](#), Ilya Dryomov)
- doc: doc/cephfs posix: remove stale information for seekdir ([pr#17658](#), "Yan, Zheng")
- doc: doc/conf.py: do not set html\_use\_smartytags explicitly ([pr#17127](#), Kefu Chai)
- doc: doc/dev: add a brief guide to serialization ([pr#20131](#), John Spray)
- doc: doc/dev/cxx: add C++11 ABI related doc ([pr#20030](#), Kefu Chai)
- doc: doc/dev/iana: document our official IANA numbers ([pr#16910](#), Sage Weil)
- doc: doc/dev/index: update rados lead ([pr#16911](#), Sage Weil)
- doc: doc/dev/macos: add doc for building on MacOS ([pr#20031](#), Kefu Chai)
- doc: doc/dev/msgr2.rst: a few notes on protocol goals ([pr#20083](#), Sage Weil)
- doc: doc/dev/perf: add doc on disabling -fomit-frame-pointer ([pr#17358](#), Kefu

Chai)

- doc: doc for mount.fuse.ceph ([issue#21539](#), [pr#19172](#), Jos Collin)
- doc: doc/man remove deprecation of ceph-disk man page title ([pr#19325](#), Alfredo Deza)
- doc: doc/mgr: Add limitations section to plugin guide ([pr#21347](#), Tim Serong)
- doc: doc/mgr: add “local pool” plugin to toc ([pr#17961](#), Kefu Chai)
- doc: doc/mgr/balancer: document ([issue#22789](#), [pr#21421](#), Sage Weil)
- doc: doc/mgr: document facilities methods using automethod directive ([pr#18680](#), Kefu Chai)
- doc: doc/mgr/plugins: add note about distinction between config and kv store ([pr#21671](#), Jan Fajerski)
- doc: doc/mgr: remove non user-facing code from doc ([pr#20372](#), Kefu Chai)
- doc: doc,os,osdc: drop and modify comments ([pr#17732](#), songweibin)
- doc: doc/rados: Add explanation of straw2 ([pr#19247](#), Shinobu Kinjo)
- doc: doc/rados/operations/bluestore-migration: document bluestore migration process ([pr#16918](#), Sage Weil)
- doc: doc/rados/operations/bluestore-migration: update docs a bit ([pr#17011](#), Sage Weil)
- doc: doc raise exceptions with a base class ([pr#18152](#), Alfredo Deza)
- doc: doc/rbd: add info for rbd group ([pr#17633](#), yonghengdexin735)
- doc: doc/rbd: add missing several commands in rbd CLI man page ([issue#14539](#), [issue#16999](#), [pr#19659](#), songweibin)
- doc: doc/rbd: correct the path of librbd python APIs ([pr#19690](#), songweibin)
- doc: doc/rbd: fix typo s/mrror/mirror ([pr#19997](#), songweibin)
- doc: doc/rbd: iSCSI Gateway Documentation ([issue#20437](#), [pr#17376](#), Aron Gunn, Jason Dillaman)
- doc: doc/rbd: specify additional ESX prerequisites ([pr#18517](#), Jason Dillaman)
- doc: doc/rbd: tweaks for the LIO iSCSI gateway ([issue#21763](#), [pr#18250](#), Jason Dillaman)
- doc: doc/rbd: tweaks to the Windows iSCSI initiator directions ([pr#18704](#), Jason Dillaman)

- doc: doc/release-notes: add jewel->kraken notes ([pr#18482](#), Sage Weil)
- doc: doc/release-notes: clarify purpose of require-osd-release ([pr#17270](#), Sage Weil)
- doc: doc/release-notes: clarify that you need to keep your existing OSD caps ([pr#18825](#), Jason Dillaman)
- doc: doc/release-notes: ensure RBD users can blacklist prior to upgrade ([issue#21353](#), [pr#17755](#), Jason Dillaman)
- doc: doc/release-notes: fix typo 'psd' to 'osd' ([pr#18695](#), wangsongbo)
- doc: doc/releases: the Kraken sleepeth, faintest sunlights flee ([pr#17424](#), Abhishek Lekshmanan)
- doc: doc/releases: update release cycle docs ([pr#18117](#), Sage Weil)
- doc: doc/rgw: add page for http frontend configuration ([issue#13523](#), [pr#20058](#), Casey Bodley)
- doc: doc/scripts: py3 compatible ([pr#17640](#), Kefu Chai)
- doc: docs: Do not use "min size = 1" as an example ([pr#17912](#), Alfredo Deza)
- doc: docs fix ceph-volume missing sub-commands ([issue#23148](#), [pr#20673](#), Alfredo Deza)
- doc: doc/start/os-recommendations.rst: bump krbd kernels ([pr#21478](#), Ilya Dryomov)
- doc: docs update ceph-deploy reference to reflect ceph-volume API ([pr#20510](#), Alfredo Deza)
- doc: document ceph-disk prepare class hierarchy ([pr#17019](#), Loic Dachary)
- doc: document include/ipaddr.h ([issue#12056](#), [pr#17613](#), Nathan Cutler)
- doc: drop duplicate line in ceph-bluestore-tool man page ([pr#19169](#), Xiaojun Liao)
- doc: eliminate useless cat statement ([pr#17154](#), Ken Dreyer)
- doc: examples: add new librbd example ([pr#18314](#), Mahati Chamarthy)
- doc: expand developer documentation of unit tests ([pr#19594](#), Nathan Cutler)
- doc: Fix a grammar error in rbd-snapshot.rst ([pr#21470](#), Zeqing Tyler Qi)
- doc: fix CFLAGS in doc/dev/cpu-profiler.rst ([pr#19752](#), Chang Liu)
- doc: fix desc of option "mon cluster log file" ([pr#18770](#), Kefu Chai)
- doc: fix doc/radosgw/admin.rst typos ([pr#17397](#), Enming Zhang)

- doc: Fix dynamic resharding doc formatting ([pr#20970](#), Ashish Singh)
- doc: fix error in osd scrub load threshold ([pr#21678](#), Dirk Sarpe)
- doc: Fixes a spelling error and a broken hyperlink ([pr#20442](#), Jordan Hus)
- doc: Fixes rbd snapshot flatten example ([issue#17723](#), [pr#17436](#), Ashish Singh)
- doc: fixes syntax in osd-config-ref ([issue#21733](#), [pr#18188](#), Joshua Schmid)
- doc: Fixes the name of the CephFS snapshot directory ([pr#18710](#), Jordan Rodgers)
- doc: fix hyper link to radosgw/config-ref ([pr#17986](#), Kefu Chai)
- doc: fix librbdpy example ([pr#20019](#), Yuan Zhou)
- doc: fix order of options in osd new ([issue#21023](#), [pr#17326](#), Neha Ojha)
- doc: fix sphinx build warnings and errors ([pr#17025](#), Alfredo Deza)
- doc: fix the desc of “osd max pg per osd hard ratio” ([pr#18373](#), Kefu Chai)
- doc: Fix typo and URL ([pr#18040](#), Jos Collin)
- doc: fix typo e.g., => e.g ([pr#18607](#), Yao Zongyou)
- doc: fix typo in bluestore-migration.rst ([pr#18389](#), Yao Zongyou)
- doc: Fix typo in mount.fuse.ceph ([pr#19215](#), Jos Collin)
- doc: fix typo in php.rst ([pr#17762](#), Yao Zongyou)
- doc: fix typo in radosgw/dynamicresharding.rst ([pr#18651](#), Alexander Ermolaev)
- doc: fix typo on specify db block device ([pr#17590](#), Xiaoxi Chen)
- doc: Fix typo s/applicatoion/application/ ([pr#20720](#), Francois Deppierraz)
- doc: Fix typos in placement-groups.rst ([pr#17973](#), Matt Boyle)
- doc: Fix typos in release notes ([pr#18950](#), Stefan Knorr)
- doc: .githubmap: Add cbodley ([pr#18946](#), Jos Collin)
- doc: githubmap: add map for GitHub contributor lookup ([pr#17457](#), Patrick Donnelly)
- doc: .githubmap, .mailmap, .organizationmap, .peoplemap: update Igor ([pr#19314](#), Igor Fedotov)
- doc: globally change CRUSH ruleset to CRUSH rule ([issue#20559](#), [pr#19435](#), Nathan Cutler)
- doc: Improved dashboard documentation ([pr#21443](#), Lenz Grimmer)

- doc: Improved hitset parameters description ([pr#19691](#), Alexey Stupnikov)
- doc: improve links in doc/releases.rst ([pr#18155](#), Nathan Cutler)
- doc: Improve mgr/restful module documentation ([pr#20717](#), Boris Ranto)
- doc: Improve the ceph fs set max\_mds command ([issue#21007](#), [pr#17044](#), Bara Ancincova)
- doc: include ceph-disk and ceph-disk-volume man pages in index ([pr#17168](#), Alfredo Deza)
- doc: init flags to 0 in rados example ([pr#20671](#), Patrick Donnelly)
- doc: Kube + Helm installation ([pr#18520](#), Alexandre Marangone)
- doc: legal: remove doc license ambiguity ([issue#23336](#), [pr#20876](#), Nathan Cutler)
- doc: lock\_timeout is a per mapping option ([pr#21563](#), Ilya Dryomov)
- doc: log-and-debug: fix default value of “log max recent” ([pr#20316](#), Nathan Cutler)
- doc: mailmap: Add Sibei, XueYu Affiliation ([pr#18395](#), Sibei Gao)
- doc: mailmap: Fixed maintenance guide URL ([pr#18076](#), Jos Collin)
- doc: mailmap, organizationmap: add Dongsheng, Liuzhong, Pengcheng, Yang Affiliation ([pr#17548](#), Dongsheng Yang)
- doc: .mailmap, .organizationmap: add Fufei, Mingqiao and Ying Affiliation ([pr#17540](#), Ying He)
- doc: .mailmap, .organizationmap: Add Liu Lei's mailmap and affiliation ([pr#17105](#), iliul)
- doc: .mailmap, .organizationmap: update JingChen, ZongyouYao, ShanchunLv's... ([pr#18960](#), Chang Liu)
- doc: mailmap: update affiliation for Mykola Golub ([pr#18069](#), Mykola Golub)
- doc: mailmap: update affiliation for Mykola Golub ([pr#19667](#), Mykola Golub)
- doc: mailmap: Update umcloud affiliation ([pr#17441](#), Yixing Yan)
- doc: make the commands in README.md properly aligned ([pr#18639](#), Yao Zongyou)
- doc/man: add “ls” to “ceph osd” command’s subcommands list ([pr#19382](#), Rishabh Dave)
- doc: “mds blacklist interval” vs manually blacklisting ([pr#18195](#), Ken Dreyer)
- doc: mgr/dashboard.rst: mention ceph.conf and ceph mgr services ([pr#20961](#), Nathan

Cutler)

- doc/mgr/plugins: mgr accessor during init causes exception ([pr#16973](#), Jan Fajerski)
- doc: mimic: doc: Updated dashboard documentation (features, SSL config) ([pr#22079](#), Lenz Grimmer)
- doc: misc fix spell errors in osd/OSD and doc ([pr#17107](#), songweibin)
- doc: misc: fix various spelling errors ([pr#20831](#), Shengjing Zhu)
- doc: Misc iSCSI doc updates ([pr#19931](#), Mike Christie)
- doc: move glance\_api\_version option to the right place ([pr#17337](#), Luo Kexue)
- doc: options.cc: document rgw config options ([pr#18007](#), Yehuda Sadeh)
- doc: organizationmap: Add Adam Wolfe Gordon's affiliation ([pr#18295](#), Adam Wolfe Gordon)
- doc: organizationmap: Add Ashish Singh affiliation ([pr#17109](#), Ashish Singh)
- doc: .organizationmap: add Xin Yuan and Yichao Li's affiliation ([pr#21170](#), Li Wang)
- doc: PendingReleaseNotes: Added note about Dashboard v2, fixed typo ([pr#21597](#), Lenz Grimmer)
- doc: PendingReleaseNotes:Announce FreeBSD availability ([pr#16782](#), Willem Jan Withagen)
- doc: PendingReleaseNotes: mention some monitor changes ([pr#21474](#), Joao Eduardo Luis)
- doc: PendingReleaseNotes: note about upmap mapping change in luminous release notes ([pr#17813](#), Sage Weil)
- doc: qa,doc: drop support of ubuntu trusty ([pr#19307](#), Kefu Chai)
- doc/rados/operations/bluestore-migration: typos and whitespace ([pr#16991](#), Sage Weil)
- doc/rados/operations/bluestore-migration: typos ([pr#17581](#), Sage Weil)
- doc: README: Improve vstart.sh usage ([pr#17644](#), Fabian Vogt)
- doc: README.md: bump up cmake to 2.8.12 ([pr#18348](#), Yan Jun)
- doc: redundant "cephfs" when set the "allow\_multimds" ([pr#20045](#), Shangzhong Zhu)
- doc: release notes: fix grammar/style nits ([pr#18876](#), Nathan Cutler)

- doc: release notes for 12.2.3 ([pr#20500](#), Abhishek Lekshmanan)
- doc: release notes for v12.1.4 Luminous ([pr#17037](#), Abhishek Lekshmanan)
- doc/release-notes: remove mention of crush weight optimization ([pr#16974](#), Sage Weil)
- doc: release-notes.rst: add Kraken v11.2.1 and update releases.rst ([pr#16879](#), Nathan Cutler)
- doc: release notes update for 10.2.10 ([pr#18148](#), Abhishek Lekshmanan)
- doc/releases: drop LTS/stable line from second chart ([pr#18153](#), Sage Weil)
- doc: Remove additional arguments when replacing OSD ([pr#18345](#), Wido den Hollander)
- doc: remove duplicated -max-buckets option desc ([pr#19737](#), Kefu Chai)
- doc: remove references to unversioned repository addresses ([pr#21357](#), Greg Farnum)
- doc: remove unused config: “osd op threads” ([pr#21319](#), Jianpeng Ma)
- doc: rename changelog with a .txt extension ([pr#18156](#), Abhishek Lekshmanan)
- doc: reorganize releases ([pr#20784](#), Abhishek Lekshmanan)
- doc: replace injectargs usage with “config set” ([pr#18789](#), John Spray)
- doc: replace region with zonegroup in configure bucket sharding section ([issue#21610](#), [pr#18063](#), Orit Wasserman)
- doc: restructure bluestore migration insructions ([pr#17603](#), Sage Weil)
- doc: Revise the Example of Bucket Policy ([pr#17362](#), zhangwen)
- doc: rgw: add a note for resharding in 12.2.1 docs ([pr#17675](#), Abhishek Lekshmanan)
- doc: rgw add some basic documentation for sync plugins & ES ([pr#15849](#), Abhishek Lekshmanan)
- doc: rgw adminops binding libraries ([pr#19164](#), hrchu)
- doc: rgw mention about tagging & bucket policies in s3api ([pr#16907](#), Abhishek Lekshmanan)
- doc: rgw: mention the civetweb support for binding to multiple ports ([issue#20942](#), [pr#17141](#), Abhishek Lekshmanan)
- doc: rm stray ”)” character from mds config ref ([pr#18228](#), Ken Dreyer)

- docs: ceph-volume CLI updates ([pr#17425](#), Alfredo Deza)
- doc: s/deamon/daemon/ ([pr#20931](#), ashitakasam)
- doc: some improvements to ceph-conf.rst ([pr#21268](#), Nathan Cutler)
- doc: Specify mount details in ceph-fuse ([pr#20071](#), Jos Collin)
- doc: SubmittingPatches: clarify PR title section ([pr#17143](#), Nathan Cutler)
- doc/templates update toctree call to include hidden entries ([pr#17076](#), Alfredo Deza)
- doc: the client inputs the pool name instead of pool ID ([pr#17672](#), Frank Yu)
- doc: typo fix ([pr#21077](#), Ashita Dashottar)
- doc: update Blacklisting and OSD epoch barrier ([issue#22542](#), [pr#19701](#), Jos Collin)
- doc: update ceph-disk with a state-transition diagram ([pr#17639](#), Kefu Chai)
- doc: update ceph iscsi kernel and package info ([pr#20020](#), Mike Christie)
- doc: Update commands and options in radosgw-admin ([pr#18267](#), Jos Collin)
- doc: update Component Technical Leads and maintainers to canonical location ([pr#18376](#), Patrick McGarry)
- doc: Update config file search paths to reflect reality ([pr#19882](#), Adam Wolfe Gordon)
- doc: updated add primary storage documentation for latest CloudStack release (4.11) ([pr#21050](#), James McClune, John Wilkins)
- doc: Update dashboard feature list (added RGW management) ([pr#21781](#), Lenz Grimmer)
- doc: updated dashboard feature list (added new RGW details, Pools) ([pr#21562](#), Lenz Grimmer)
- doc: Updated dashboard feature list ([pr#21693](#), Lenz Grimmer)
- doc: Updated dashboard v2 feature list ([pr#20755](#), Lenz Grimmer)
- doc: Updated documentation for Zabbix Mgr module ([pr#18356](#), Wido den Hollander)
- doc: update default value of option mon\_sync\_timeout ([pr#17802](#), Yao Guotao)
- doc: update default value of parameter mon\_subscribe\_interval ([pr#17669](#), yaoguotao)
- doc: Update docs to remove gitbuilder and add shaman references ([pr#17022](#),

Alfredo Deza)

- doc: updated the dashboard feature list ([pr#21531](#), Lenz Grimmer)
- doc: Updated the get-packages.rst to luminous ([pr#20815](#), Kai Wagner)
- doc: update firewall doc to mention ceph-mgr ([pr#17974](#), John Spray)
- doc: update iSCSI upstream kernel to 4.16 ([pr#20695](#), Mike Christie)
- doc: update link to placing-different-pools ([pr#17833](#), Mohamad Gebai)
- doc: update Li Wang Affiliation ([pr#18060](#), Li Wang)
- doc: update man page to explain ceph-volume support bluestore ([issue#22663](#), [pr#19960](#), lijing)
- doc: Update manual deployment ([issue#20309](#), [pr#15811](#), Jens Rosenboom)
- doc: update mgr/dashboard doc about standbys ([pr#19879](#), John Spray)
- doc: Update mgr doc on how to enable Zabbix module ([pr#16861](#), Wido den Hollander)
- doc: update mgr related auth settings ([pr#20126](#), Kefu Chai)
- doc: Update monitoring.rst ([pr#20630](#), Jos Collin)
- doc: update rbd-mirroring documentation ([issue#20701](#), [pr#16908](#), Jason Dillaman)
- doc: update references to use ceph-volume ([pr#19241](#), Alfredo Deza)
- doc: update releases to the current state ([pr#17364](#), Abhishek Lekshmanan)
- doc: Updates to bluestore migration doc ([pr#17602](#), David Galloway)
- doc: v12.2.5 luminous release notes ([pr#21621](#), Abhishek Lekshmanan)
- doc: various cleanups ([pr#18480](#), Kefu Chai)
- examples: fix link order in librados example Makefile ([pr#17842](#), Mahati Chamopathy)
- Fix ceph-mgr restarts ([pr#22051](#), Boris Ranto)
- follow-up fixups for atomic\_t spinlocks ([pr#17611](#), Jesse Williamson)
- githubmap: Add ktdreyer ([pr#19209](#), Jos Collin)
- include/buffer.h: fix typo in comment ([pr#17489](#), mychoxin)
- include/ceph\_features: fix OS\_PERF\_STAT\_NS's incarnation ([pr#21467](#), Kefu Chai)
- install-deps.sh: fix an error condition expression ([pr#20819](#), Yao Guotao)

- java/native: fix milliseconds to mtime/atime conversion ([pr#17460](#), dengquan)
- java/native: s/jni: lstat/jni: stat in native\_ceph\_stat ([pr#20142](#), Shangzhong Zhu)
- KStore: statfs needs extra includes on FreeBSD ([pr#21429](#), Willem Jan Withagen)
- kv/leveldb: fix deadlock when close db ([pr#16643](#), Zengran)
- kv: unify {create\_and\_,}open() methods ([pr#18177](#), Kefu Chai)
- librados: add async interfaces for use with Networking TS ([pr#19054](#), Casey Bodley)
- librados: block MgrClient::start\_command until mgrmap ([pr#21832](#), John Spray, Kefu Chai)
- librados: extend C API for so it accepts keys with NUL chars ([pr#20314](#), Piotr Dałek)
- librados: Fix a potential risk of buffer::list::claim\_prepnd(list& b... ([issue#21338](#), [pr#17661](#), Guan yunfei))
- librados: fix potential race condition if notify immediately fails ([issue#23966](#), [pr#21859](#), Jason Dillaman)
- librados: getter for min compatible client versions ([pr#20080](#), Jason Dillaman)
- librados: invalid free() in rados\_getxattrs\_next() ([issue#22042](#), [pr#20260](#), Gu Zhongyan)
- librados: make OPERATION\_FULL\_FORCE the default for rados\_remove() ([issue#22413](#), [pr#20534](#), Kefu Chai)
- librbd: abstract hard-coded journal and cache hooks on IO path ([pr#20682](#), Jason Dillaman)
- librbd: Add a function to list image watchers ([pr#19188](#), Adam Wolfe Gordon)
- librbd: add API function to get image name ([pr#20935](#), Mykola Golub)
- librbd: added preprocessor macro for detecting compare-and-write support ([issue#22036](#), [pr#18708](#), Jason Dillaman)
- librbd: add eventtrace support ([pr#19251](#), Mahati Chamopathy)
- librbd: add preliminary support for new operation feature bit ([pr#19903](#), Jason Dillaman)
- librbd: address coverity false positives ([pr#17696](#), Amit Kumar)
- librbd: address coverity false positives ([pr#17721](#), Amit Kumar)

- librbd: auto-remove trash snapshots when image is deleted ([issue#22873](#), [pr#20376](#), Jason Dillaman)
- librbd: by default use new format for deep copy destination ([pr#20222](#), Mykola Golub)
- librbd: cache last index position to accelerate snap create/rm ([issue#22716](#), [pr#19974](#), Song Shun)
- librbd: cannot clone all image-metas if we have more than 64 key/value pairs ([pr#18327](#), PCzhangPC)
- librbd: cannot copy all image-metas if we have more than 64 key/value pairs ([pr#18328](#), PCzhangPC)
- librbd: clean up ManagedLock log prefix ([pr#20159](#), shun-s)
- librbd: compare and write against a clone can result in failure ([issue#20789](#), [pr#18887](#), Jason Dillaman)
- librbd: deep\_copy: don't create snapshots above snap\_id\_end ([pr#19383](#), Mykola Golub)
- librbd: default localize parent reads to false ([issue#20941](#), [pr#16882](#), Jason Dillaman)
- librbd: default to sparse-reads for any IO operation over 64K ([issue#21849](#), [pr#18405](#), Jason Dillaman)
- librbd: disable ENOENT tracking within the object cacher ([issue#23597](#), [pr#21308](#), Jason Dillaman)
- librbd: disallow creation of v1 image format ([pr#20460](#), Julien COLLET, Julien Collet)
- librbd: don't read metadata twice on image open ([pr#18542](#), Mykola Golub)
- librbd: drop redundant check for null ImageCtx ([pr#18265](#), Jianpeng Ma)
- librbd: filter out potential race with image rename ([issue#18435](#), [pr#19618](#), Jason Dillaman)
- librbd: fix coverity warning for uninitialized member ([pr#18129](#), Li Wang)
- librbd: fix deep copy a child-image ([pr#20099](#), songweibin)
- librbd: fix don't send get\_stripe\_unit\_count if striping is not enabled ([issue#21360](#), [pr#17660](#), Yanhu Cao)
- librbd: fix issues discovered in clone v2 during upgrade tests ([issue#22979](#), [pr#20406](#), Jason Dillaman)

- librbd: fix missing return in NotifyMessage::get\_notify\_op ([pr#20656](#), Yao Zongyou)
- librbd: fix rbd close race with rewatch ([pr#21141](#), Song Shun)
- librbd: fix refuse to release lock when cookie is the same at rewatch ([pr#20868](#), Song Shun)
- librbd: fix structure size check in rbd\_mirror\_image\_get\_info/status ([pr#20478](#), Mykola Golub)
- librbd: force removal of a snapshot cannot ignore dependent children ([issue#22791](#), [pr#20105](#), Jason Dillaman)
- librbd: generalized deep copy function ([pr#16238](#), Mykola Golub)
- librbd: group and snapshot cleanup ([pr#19990](#), Jason Dillaman)
- librbd: group snapshots ([pr#11544](#), Victor Denisov, Jason Dillaman)
- librbd: hold cache\_lock while clearing cache nonexistence flags ([issue#21558](#), [pr#17992](#), Jason Dillaman)
- librbd: image-meta config overrides should be dynamically refreshed ([issue#21529](#), [pr#18042](#), Dongsheng Yang, Jason Dillaman)
- librbd: initial hooks for clone v2 support ([pr#20176](#), Jason Dillaman)
- librbd: initialization of state member variables ([pr#16866](#), amitkuma)
- librbd: Initializing members image,operation,journal ([pr#16934](#), amitkuma)
- librbd: Initializing member variables ([pr#16867](#), amitkuma)
- librbd: journal should ignore -EILSEQ errors from compare-and-write ([issue#21628](#), [pr#18099](#), Jason Dillaman)
- librbd,librados: do not include stdbool.h in C++ headers ([pr#19945](#), Kefu Chai)
- librbd: list\_children should not attempt to refresh image ([issue#21670](#), [pr#18114](#), Jason Dillaman)
- librbd: minor cleanup of the IO pathway ([pr#20560](#), Jason Dillaman)
- librbd: minor code cleanup ([pr#21165](#), songweibin)
- librbd: missing 'return' in deep\_copy::ObjectCopyRequest::send\_read\_object ([pr#21493](#), Mykola Golub)
- librbd: new tag should use on-disk committed position ([issue#22945](#), [pr#20423](#), Jason Dillaman)

- librbd: object map batch update might cause OSD suicide timeout ([issue#21797](#), [pr#18315](#), Jason Dillaman)
- librbd: possible deadlock with synchronous maintenance operations ([issue#22120](#), [pr#18909](#), Jason Dillaman)
- librbd: potential crash if object map check encounters error ([issue#22819](#), [pr#20214](#), Jason Dillaman)
- librbd: potential race between discard and writeback ([pr#21248](#), Jason Dillaman)
- librbd: potential race in RewatchRequest when resetting watch\_handle ([pr#20420](#), Mykola Golub)
- librbd: prefer templates to macros ([pr#19912](#), Adam C. Emerson)
- librbd: prevent overflow of discard API result code ([issue#21966](#), [pr#18923](#), Jason Dillaman)
- librbd: prevent watcher from unregistering with in-flight actions ([issue#23955](#), [pr#21763](#), Jason Dillaman)
- librbd: refresh image after applying new metadata ([issue#21711](#), [pr#18158](#), Jason Dillaman)
- librbd: release lock executing deep copy progress callback ([issue#23929](#), [pr#21727](#), Mykola Golub)
- librbd: remove unused member in FlattenRequest ([pr#19416](#), Mykola Golub)
- librbd: remove unused variables from ReadResult refactor ([pr#18277](#), Jason Dillaman)
- librbd: rename of non-existent image results in seg fault ([issue#21248](#), [pr#17502](#), Jason Dillaman)
- librbd: set deleted parent pointer to null ([issue#22158](#), [pr#19003](#), Jason Dillaman)
- librbd: should not set self as remote peer ([pr#17300](#), songweibin)
- librbd: small cleanup for recently merged code ([pr#20578](#), Mykola Golub)
- librbd: snapshots should be created/removed against data pool ([issue#21567](#), [pr#18043](#), Jason Dillaman)
- librbd: speed up object map disk usage and resize ([pr#20218](#), shun-s)
- librbd: speed up sparse copy when object map is available ([pr#18967](#), Song Shun)
- librbd: update mirror::EnableRequest diagram according to code ([pr#19130](#), Mykola Golub)

- librbd: use steady clock to measure elapsed time in AioCompletion ([pr#20007](#), Mohamad Gebai)
- librbd: validate if dst group snap name is the same with src ([pr#20395](#), songweibin)
- log: Fix AddressSanitizer: new-delete-type-mismatch ([issue#23324](#), [pr#20930](#), Brad Hubbard)
- log: fix build on osx ([pr#18213](#), Kefu Chai)
- log: silence warning from -Wsign-compare ([pr#18326](#), Jos Collin)
- log: Use the coarse real time clock in log timestamps ([pr#18141](#), Adam C. Emerson)
- mds: check metadata pool not cluster is full ([issue#22483](#), [pr#19602](#), Patrick Donnelly)
- mds: fix CEPH\_STAT\_RSTAT definition ([pr#21633](#), "Yan, Zheng")
- mds: get rid of the "if" check which is unnecessary inside a loop ([pr#18904](#), dongdong tao)
- mds: Remove redundant null pointer check ([pr#19750](#), Brad Hubbard)
- mds: simplify the code logic in replay\_alloc\_ids ([pr#18893](#), dongdong tao)
- mempool: fix lack of pool names in mempool:dump output for JSON format ([pr#18329](#), Igor Fedotov)
- messages: Initialization of uninitialized members various classes ([pr#16848](#), amitkuma)
- messages/MDentryLink: add const to member function ([pr#15479](#), yonghengdexin735)
- messages,test,msg: initialize h,reply\_type,owner ([pr#17767](#), Amit Kumar)
- mgr: add mgr daemon to DaemonStateIndex with metadata (hostname) ([issue#23286](#), [pr#20875](#), Jan Fajerski)
- mgr: add missing call to pick\_addresses ([issue#20955](#), [pr#16940](#), John Spray)
- mgr: add the ip addr of standbys ([pr#16476](#), huanwen ren)
- mgr: add units to performance counters ([issue#22747](#), [pr#20152](#), Rubab Syed)
- mgr: allow service daemons to unregister from ServiceMap ([pr#20761](#), Sage Weil)
- mgr: apply a threshold to perf counter prios ([pr#16699](#), John Spray)
- mgr: balancer: fixed mistype "AttributeError: 'Logger' object has no attribute 'err'" ([pr#20130](#), Konstantin Shalygin)

- mgr: centralized setting/getting of mgr configs ([pr#21442](#), John Spray, Rubab Syed)
- mgr: ceph-mgr: can not change prometheus port for mgr ([pr#17746](#), wujian)
- mgr: common interface for TSDB modules ([pr#17735](#), Jan Fajerski, John Spray, My Do)
- mgr/dashboard: Adapt help text if server\_addr is not set ([pr#21640](#), Volker Theile)
- mgr/dashboard: Adapt RBD form to new application\_metadata type ([pr#21602](#), Volker Theile)
- mgr/dashboard: Add Api module ([pr#21126](#), Tiago Melo)
- mgr/dashboard: Add 'autofocus' directive ([pr#21559](#), Volker Theile)
- mgr/dashboard: Add CdDatePipe ([pr#21087](#), Ricardo Marques)
- mgr/dashboard: Add 'cd-error-panel' component to display error messages ([pr#21558](#), Volker Theile)
- mgr/dashboard: Add 'cd-loading-panel' component ([pr#21618](#), Volker Theile)
- mgr/dashboard: Add custom validators ([pr#21041](#), Volker Theile)
- mgr/dashboard: Add DimlessBinaryDirective ([pr#20972](#), Ricardo Marques)
- mgr/dashboard: Add ErasureCodeProfile controller ([issue#23345](#), [pr#20920](#), Sebastian Wagner, Stephan Müller)
- mgr/dashboard: Add 'forceIdentifier' attribute to datatable ([pr#21497](#), Volker Theile)
- mgr/dashboard: Add helper component ([pr#20971](#), Ricardo Marques)
- mgr/dashboard: additional fixes to block pages ([pr#20941](#), Jason Dillaman)
- mgr/dashboard: Add minimalistic browsable API ([pr#20873](#), Sebastian Wagner)
- mgr/dashboard: Add notification service/component ([pr#21078](#), Tiago Melo)
- mgr/dashboard: Add Pool-create to the backend ([issue#23345](#), [pr#20865](#), Sebastian Wagner)
- mgr/dashboard: Add RGW user and bucket management features ([pr#21351](#), Volker Theile)
- mgr/dashboard: Adds reusable deletion dialog ([pr#20899](#), Stephan Müller, Tiago Melo)

- mgr/dashboard: Add submit button component ([pr#21011](#), Tiago Melo)
- mgr/dashboard: Add usage bar component ([pr#21128](#), Ricardo Marques)
- mgr/dashboard: Angular modules cleanup ([pr#21402](#), Tiago Melo)
- mgr/dashboard: Asynchronous tasks (frontend) ([pr#20962](#), Ricardo Marques)
- mgr/dashboard: awsauth: fix python3 string decode problem ([pr#21875](#), Ricardo Dias)
- mgr/dashboard: Change font-family of checkbox ([pr#21787](#), Tiago Melo)
- mgr/dashboard: Clean up Pylint warnings ([pr#21694](#), Sebastian Wagner)
- mgr/dashboard: Convert floating values to bytes ([pr#21677](#), Stephan Müller)
- mgr/dashboard: Convert the RBD feature names to a list of strings ([pr#21024](#), Tatjana Dehler)
- mgr/dashboard: Deletion dialog falsely executes deletion when pressing 'Cancel' ([pr#22032](#), Volker Theile)
- mgr/dashboard: Display notification if RGW is not configured ([pr#21977](#), Volker Theile)
- mgr/dashboard: Display RBD form errors on submission ([pr#21529](#), Ricardo Marques)
- mgr/dashboard: Enable object rendering in KV-table ([pr#21701](#), Stephan Müller)
- mgr/dashboard: fix 500 error on block device iSCSI status page ([pr#20928](#), Jason Dillaman)
- mgr/dashboard: fix dashboard python 3 support ([pr#21007](#), Ricardo Dias)
- mgr/dashboard: Fix data race and use-before-assignment ([pr#21590](#), Sebastian Wagner)
- mgr/dashboard: fixed password generation in Auth controller ([issue#23404](#), [pr#21006](#), Ricardo Dias)
- mgr/dashboard: Fixes documentation link- to open in new tab ([pr#22262](#), Kanika Murarka)
- mgr/dashboard: Fixes type error in RBD form ([pr#21681](#), Stephan Müller)
- mgr/dashboard: fix frontend e2e tests ([pr#20943](#), Tiago Melo)
- mgr/dashboard: fix FS status on old MDS daemons ([issue#20692](#), [pr#16960](#), John Spray)
- mgr/dashboard: fix linting problem ([pr#22277](#), Tiago Melo)

- mgr/dashboard: Fix missing \$event on deletion modal ([pr#21667](#), Ricardo Marques)
- mgr/dashboard: Fix moment.js deprecation warning ([pr#22052](#), Tiago Melo)
- mgr/dashboard: Fix objects named default are inaccessible ([pr#20976](#), Sebastian Wagner)
- mgr/dashboard: Fix RBD task metadata ([pr#22152](#), Tiago Melo)
- mgr/dashboard: Fix table without fetchData ([pr#21086](#), Ricardo Marques)
- mgr/dashboard: Fix the data table action selector ([pr#21270](#), Stephan Müller)
- mgr/dashboard: fix two type errors found by mypy ([pr#21774](#), Sebastian Wagner)
- mgr/dashboard: Handle errors during deletion ([pr#22029](#), Volker Theile)
- mgr/dashboard: Implement a RGW proxy ([pr#21258](#), Volker Theile, Patrick Nawracay)
- mgr/dashboard: Improve background tasks style ([pr#21462](#), Ricardo Marques)
- mgr/dashboard: improve error handling ([pr#18182](#), Nick Erdmann)
- mgr/dashboard: Improve error panel ([pr#21978](#), Volker Theile)
- mgr/dashboard: Improve npm start script ([pr#20989](#), Ricardo Marques)
- mgr/dashboard: Improve table search ([pr#20807](#), Stephan Müller)
- mgr/dashboard: Load the datatable content on component initialization ([pr#21595](#), Volker Theile)

- mgr/dashboard: Navbar dropdown button does not respond for mobile browsers ([pr#21979](#), Volker Theile)
- mgr/dashboard: Notification improvements ([pr#21350](#), Tiago Melo)
- mgr/dashboard: pool: fix python3 dict\_keys error ([pr#21636](#), Ricardo Dias)
- mgr/dashboard: Pool listing ([pr#21353](#), Stephan Müller)
- mgr/dashboard: rbd: add @AuthRequired to snapshots controller ([pr#21517](#), Ricardo Dias)
- mgr/dashboard: RBD copy, RBD flatten and snapshot clone (frontend) ([pr#21526](#), Ricardo Marques, Ricardo Dias)
- mgr/dashboard: RBD management (frontend) ([pr#21385](#), Ricardo Marques)
- mgr/dashboard: Refactor multiple duplicates of get\_rate() ([pr#21022](#), Sebastian Wagner)
- mgr/dashboard: Refactor RGW backend ([pr#21855](#), Volker Theile)
- mgr/dashboard: Rename and refactor ApiInterceptorService class ([pr#21386](#), Volker Theile)
- mgr/dashboard: Replace font-awesome with fork-awesome ([pr#21327](#), Lenz Grimmer)
- mgr/dashboard: restcontroller: fix detection of id args in element requests ([pr#21290](#), Ricardo Dias)
- mgr/dashboard: RESTController improvements ([pr#21516](#), Ricardo Dias)
- mgr/dashboard: run-tox: pass CEPH\_BUILD\_DIR value into tox script ([pr#21445](#), Ricardo Dias)
- mgr: dashboard: show per pool IOPS on health page (#22495) ([issue#22495](#), [pr#19981](#), Konstantin Shalygin)
- mgr/dashboard: Support aditional info on 'cd-view-cache' ([pr#21060](#), Ricardo Marques)
- mgr/dashboard: TaskManager bug fixes ([pr#21240](#), Ricardo Dias)
- mgr/dashboard: Update selected items on table refresh ([pr#21099](#), Ricardo Marques)
- mgr/dashboard: Use Bootstrap CSS ([pr#21780](#), Volker Theile)
- mgr/dashboard: using RoutesDispatcher as HTTP request dispatcher ([pr#21239](#), Ricardo Dias)
- mgr/dashboard\_v2: add mgr to the list of perf counters ([pr#20783](#), Tiago Melo)

- mgr/dashboard\_v2: add mocked service provider for TcmuIscsiService ([pr#20775](#), Tiago Melo)
- mgr/dashboard\_v2: Add toggle able columns ([pr#20806](#), Stephan Müller)
- mgr/dashboard\_v2: Configuration settings support ([pr#20743](#), Ricardo Dias)
- mgr/dashboard\_v2: fix and improve table details ([pr#20811](#), Tiago Melo)
- mgr/dashboard\_v2: Fix cephfs template table usage ([pr#20804](#), Stephan Müller)
- mgr/dashboard\_v2: fix cluster configuration page ([pr#20821](#), Tiago Melo)
- mgr/dashboard\_v2: Improve charts tooltips ([pr#20757](#), Tiago Melo)
- mgr/dashboard\_v2: Pool controller ([pr#20823](#), Ricardo Dias)
- mgr/dashboard\_v2: Rotate the refresh icon on load ([pr#20805](#), Stephan Müller)
- mgr: die on bind() failure ([pr#20595](#), John Spray)
- mgr: disconnect unregistered service daemon when report received ([issue#22286](#), [pr#19261](#), Jason Dillaman)
- mgr: emit cluster log message on serve() exception ([issue#21999](#), [pr#18672](#), John Spray)
- mgr: Expose rgw perf counters ([pr#21269](#), Boris Ranto)
- mgr: fix "access denied" message ([pr#19518](#), John Spray)
- mgr: fix crashable DaemonStateIndex::get calls ([issue#17737](#), [pr#17933](#), John Spray)
- mgr: fix crash in MonCommandCompletion ([issue#21157](#), [pr#17308](#), John Spray)
- mgr: fixes python error handling ([issue#23406](#), [pr#21005](#), Ricardo Dias)
- mgr: fix MSG\_MGR\_MAP handling ([pr#20892](#), Gu Zhongyan)
- mgr: fix "osd status" command exception if OSD not in pgmap stats ([issue#21707](#), [pr#18173](#), Yanhu Cao)
- mgr: fix py3 support ([issue#22880](#), [pr#20362](#), Kefu Chai)
- mgr: fix py calls for dne service perf counters ([issue#21253](#), [pr#17605](#), John Spray)
- mgr: implement completion of osd MetadataUpdate ([issue#21159](#), [pr#16925](#), Yanhu Cao)
- mgr: implement 'osd safe-to-destroy' and 'osd ok-to-stop' commands ([pr#16976](#), Sage Weil)

- mgr: improved module loading for error reporting etc ([issue#21999](#), [issue#21683](#), [issue#21502](#), [pr#19235](#), John Spray)
- mgr: improve reporting on unloadable modules ([issue#23358](#), [pr#20921](#), John Spray)
- mgr: increase time resolution of Commit/Apply OSD latencies ([pr#19232](#), Коренберг Марк)
- mgr: initialize PyModuleRegistry sooner ([issue#22918](#), [pr#20321](#), John Spray)
- mgr: In plugins 'module' classes need not to be called "Module" anymore ([issue#17454](#), [pr#18526](#), Kefu Chai, bhavishyagopesh)
- mgr: locking fixes ([issue#21158](#), [pr#17309](#), John Spray)
- mgr: mgr/balancer: cast config vals to int or float ([issue#22429](#), [pr#19493](#), Dan van der Ster)
- mgr: mgr/balancer: don't use 'foo' tags on commands ([issue#22361](#), [pr#19482](#), John Spray)
- mgr: mgr/balancer: fix KeyError in balancer rm ([issue#22470](#), [pr#19578](#), Dan van der Ster)
- mgr: mgr/balancer: fix OPTIONS definition ([pr#21620](#), John Spray)
- mgr: mgr/balancer: fix upmap; default balancer module enabled ([pr#18691](#), Sage Weil)
- mgr: mgr/balancer: make crush-compat mode work ([pr#17983](#), Sage Weil)
- mgr: mgr/balancer: mgr module to automatically balance PGs across OSDs ([pr#16272](#), Spandan Kumar Sahu, Sage Weil)
- mgr: mgr/balancer: more pool-specific enhancements ([pr#20225](#), xie xingguo)
- mgr: mgr/balancer: pool-specific optimization support and bug fixes ([pr#20154](#), xie xingguo)
- mgr: mgr/balancer: replace magic value of -1 for DEFAULT\_CHOOSE\_ARGS ([pr#20258](#), Kefu Chai)
- mgr: mgr/balancer: skip CRUSH\_ITEM\_NONE ([pr#18894](#), Sage Weil)
- mgr: mgr/balancer: two more fixes ([pr#20180](#), xie xingguo)
- mgr: mgrc: free MMgrClose in handle\_mgr\_close ([issue#23846](#), [pr#21626](#), Casey Bodley)
- mgr: mgr/DaemonServer: add overrides value to 'config show' ([pr#21093](#), Gu Zhongyan)

- mgr: mgr/DaemonServer.cc: [Cleanup] Change to using get\_val template function ([pr#18717](#), Shinobu Kinjo)
- mgr: mgr/DaemonServer: [Cleanup] Remove redundant code ([pr#18716](#), Shinobu Kinjo)
- mgr: mgr/dashboard: add configuration setting browser ([issue#22522](#), [pr#20043](#), Rubab Syed)
- mgr: mgr/dashboard: add image id to mgr rbd info instead of block\_name\_prefix ([pr#20884](#), zouaiguo)
- mgr: mgr/dashboard: Add monitor list ([pr#19632](#), Rubab Syed)
- mgr: mgr/dashboard: Add RGW user and bucket lists (read-only) ([pr#20869](#), Volker Theile)
- mgr: mgr/dashboard: add TLS ([pr#21627](#), John Spray)
- mgr: mgr/dashboard: Add toBytes() method to FormatterService ([pr#20978](#), Volker Theile)
- mgr: mgr/dashboard: asynchronous task support ([pr#20870](#), Ricardo Dias)
- mgr: mgr/dashboard: change raw usage chart's color depending on usage ([pr#17421](#), Nick Erdmann)
- mgr: mgr/dashboard: fix audit log loading ([pr#18848](#), John Spray)
- mgr: mgr/dashboard: Fix backend tests for newer CherryPy versions ([pr#20778](#), Patrick Nawracay)
- mgr: mgr/dashboard: Fix PG status coloring ([pr#19431](#), Wido den Hollander)
- mgr: mgr/dashboard: format tooltip's label as user friendly string ([pr#18769](#), Yao Zongyou)
- mgr: mgr/dashboard: handle null in format\_number ([issue#21570](#), [pr#17991](#), John Spray)
- mgr: mgr/dashboard: HTTP request logging ([pr#20797](#), Ricardo Dias)
- mgr: mgr/dashboard: Improve auth interceptor ([pr#20847](#), Volker Theile)
- mgr: mgr/dashboard: performance counter browsers ([issue#22521](#), [pr#19922](#), Rubab-Syed)
- mgr: mgr/dashboard: RBD management (backend) ([pr#21360](#), Ricardo Dias)
- mgr: mgr/dashboard: Remove unused code ([pr#21045](#), Volker Theile)
- mgr: mgr/dashboard: Remove useless code ([pr#20958](#), Volker Theile)

- mgr: mgr/dashboard: show warnings if data is out of date or mons are down ([pr#18847](#), John Spray)
- mgr: mgr/dashboard: sort servers and OSDs in OSD list ([issue#21572](#), [pr#17993](#), John Spray)
- mgr: mgr/dashboard: use rel="icon" for favicon ([pr#18013](#), Kefu Chai)
- mgr: mgr/dashboard v2: Add CSS class for required form fields ([pr#20747](#), Volker Theile)
- mgr: mgr/dashboard\_v2: Add RBD create functionality to the Python backend ([pr#20751](#), Tatjana Dehler)
- mgr: mgr/dashboard v2: Add units to performance counters ([pr#20742](#), Volker Theile)
- mgr: mgr/dashboard v2: Display loading indicator in datatables during first load ([pr#20744](#), Volker Theile)
- mgr: mgr/dashboard v2: Don't show details if multiple OSDs are selected ([pr#20772](#), Volker Theile)
- mgr: mgr/dashboard v2: implement can\_run method ([pr#20728](#), John Spray)
- mgr: mgr/dashboard\_v2: Initial submission of a web-based management UI (replacement for the existing dashboard) ([pr#20103](#), Stephan Müller, Lenz Grimmer, Tiago Melo, Ricardo Marques, Sebastian Wagner, Patrick Nawracay, Ricardo Dias, Volker Theile, Kai Wagner, Tatjana Dehler)
- mgr: mgr/dashboard v2: Introduce CdTableSelection model ([pr#20746](#), Volker Theile)
- mgr: mgr/dashboard\_v2: Removed unused tools.detail\_route() ([pr#20765](#), Sebastian Wagner)
- mgr: mgr/influx: Added Additional Stats ([pr#21424](#), mhdo2)
- mgr: mgr/influx: Add InfluxDB SSL Option ([pr#19374](#), Tobias Gall)
- mgr: mgr/influx: Only split string on first occurrence of dot (.) ([issue#23996](#), [pr#21795](#), Wido den Hollander)
- mgr: mgr/influx: PEP-8 and other fixes to Influx module ([pr#19229](#), Wido den Hollander)
- mgr: mgr/influx: Various fixes and improvements ([pr#20187](#), Wido den Hollander)
- mgr: mgr/influx: Various time fixes ([pr#20494](#), Wido den Hollander)
- mgr: mgr/localpool: default to 3x; allow min\_size adjustment ([pr#18089](#), Sage Weil)

- mgr: mgr/MgrClient: guard send\_pgstats() with lock ([issue#23370](#), [pr#20909](#), Kefu Chai)
- mgr: mgr/MgrClient: service registration filtered by service name instead of daemon name ([pr#21459](#), runsisi)
- mgr: mgr/PGMap: drop REQUEST\_{SLOW,STUCK} HEALTH\_WARNS ([pr#19114](#), Kefu Chai)
- mgr: mgr/prometheus: add ceph\_disk\_occupation series ([issue#21594](#), [pr#18021](#), John Spray)
- mgr: mgr/prometheus: add missing ‘deep’ state to PG\_STATES in ceph-mgr prometheus plugin ([issue#22116](#), [pr#18890](#), Peter Woodman)
- mgr: mgr/prometheus: Fix for MDS metrics ([issue#20899](#), [pr#17318](#), John Spray, Jeremy H Austin)
- mgr: mgr/prometheus: fix PG state names ([pr#21288](#), John Spray)
- mgr: mgr/prometheus: Skip bogus entries ([pr#20456](#), Boris Ranto)
- mgr: mgr/prometheus: skip OSD output if missing from CRUSH devices ([pr#20644](#), John Spray)
- mgr: mgr/restful: A couple of restful fixes ([pr#18649](#), Boris Ranto)
- mgr: mgr/restful: cleaner message when not configured ([issue#21292](#), [pr#17573](#), John Spray)
- mgr: mgr/smart: fix python3 module loading ([pr#21047](#), Ricardo Dias)
- mgr: mgr/status: fix ceph fs status returns error ([issue#21752](#), [pr#18233](#), Yanhu Cao)
- mgr: mgr/status: format byte quantities in base 2 multiples ([issue#21189](#), [pr#17380](#), John Spray)
- mgr: mgr/telemetry: Add Ceph Telemetry module to send reports back to project ([pr#21970](#), Wido den Hollander)
- mgr: mgr/zabbix: fix div by zero ([issue#21518](#), [pr#17931](#), John Spray)
- mgr: mgr/zabbix: ignore osd with 0 kb capacity ([issue#21904](#), [pr#18809](#), Ilja Slepnev)
- mgr: mgr/zabbix: Implement health checks ([pr#20198](#), Wido den Hollander)
- mgr: mgr/zabbix: Send max, min and avg PGs of OSDs to Zabbix ([pr#21043](#), Wido den Hollander)
- mgr: mgr/Zabbix: Various fixes to Zabbix module ([pr#19452](#), Wido den Hollander)

- mgr: mimic: mgr/telegraf: Telegraf module for Ceph Mgr ([pr#22013](#), Wido den Hollander)
- mgr: Modify mgr-influx module database check to not require admin privileges ([pr#18102](#), Benjeman Meekhof)
- mgr: mon,mgr: improve ‘mgr module disable’ cmd ([pr#21188](#), Gu Zhongyan)
- mgr: mon, mgr: move “osd pool stats” command to mgr and mgr python module ([pr#19985](#), Chang Liu)
- mgr: mon/MgrStatMonitor: fix formatting of pending\_digest ([issue#22991](#), [pr#20426](#), Patrick Donnelly)
- mgr,mon: mon/MgrMonitor: read cmd desc if empty on update\_from\_paxos() ([issue#21300](#), [pr#17846](#), Joao Eduardo Luis)
- mgr,mon: mon,mgr: remove single wildcard ‘\*’ from ceph comand line description ([pr#21139](#), Gu Zhongyan)
- mgr,mon: mon/mgr: sync “mgr\_command\_descs”, “osd\_metadata” and “mgr\_metadata” prefixes to new mons ([issue#21527](#), [pr#17929](#), huanwen ren)
- mgr,mon: mon/MonCommands: mgr metadata - improve parameter naming consistency ([issue#23330](#), [pr#20866](#), Jan Fajerski)
- mgr: preventing blank hostname in DaemonState ([issue#20887](#), [issue#21060](#), [pr#17138](#), liuchang0812)
- mgr: prometheus: added osd commit/apply latency metrics (#22718) ([issue#22718](#), [pr#19980](#), Konstantin Shalygin)
- mgr: prometheus: Don’t crash on OSDs without metadata ([pr#20539](#), Christopher Blum)
- mgr: prometheus fix metadata labels ([pr#21557](#), Jan Fajerski)
- mgr: prometheus: set metadata metrics value to ‘1’ (#22717) ([issue#22717](#), [pr#19979](#), Konstantin Shalygin)
- mgr: pybind/mgr/balancer: add sanity check against empty adjusted\_map ([pr#20836](#), xie xingguo)
- mgr: pybind/mgr/balancer: fix pool-deletion vs auto-optimization race ([pr#20706](#), xie xingguo)
- mgr: pybind/mgr/balancer: fix sanity check against empty weight-set ([pr#20278](#), xie xingguo)
- mgr: pybind/mgr/balancer: increase bad\_steps properly ([pr#20194](#), xie xingguo)
- mgr: pybind/mgr/balancer: load weight-set from ms ([pr#20197](#), xie xingguo)

- mgr: pybind/mgr/balancer: more specific command outputs ([pr#20305](#), xie xingguo)
- mgr: pybind/mgr/balancer: remove optimization plan properly ([pr#20224](#), xie xingguo)
- mgr: pybind/mgr/balancer: two more fixes ([pr#20788](#), xie xingguo)
- mgr: pybind/mgr/dashboard: add url\_prefix ([issue#20568](#), [pr#17119](#), Nick Erdmann)
- mgr: pybind/mgr/dashboard: fix duplicated slash in html href ([issue#22851](#), [pr#20229](#), Shengjing Zhu)
- mgr,pybind: mgr/dashboard: fix pool size base conversion ([pr#16771](#), Yixing Yan)
- mgr: pybind/mgr/dashboard: fix reverse proxy support ([issue#22557](#), [pr#19758](#), Nick Erdmann)
- mgr,pybind: mgr/iostat: print output as a table ([pr#21338](#), Mohamad Gebai)
- mgr: pybind/mgr/localpool: module to automagically create localized pools ([pr#17528](#), Sage Weil)
- mgr: pybind/mgr/mgr\_module: add default param for MgrStandbyModule.get\_con... ([pr#19948](#), Kefu Chai)
- mgr: pybind/mgr/mgr\_module: make rados handle available to all modules ([pr#19972](#), Sage Weil)
- mgr: pybind/mgr\_module: move PRIO\_\* and PERFCOUNTER\_\* to MgrModule class ([pr#18251](#), Jan Fajerski)
- mgr: pybind/mgr: new 'hello world' mgr module skeleton ([pr#19491](#), Yaarit Hatuka)
- mgr: pybind/mgr/prometheus: add file\_sd\_config command ([pr#21061](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: add osd\_in/out metric; make osd\_weight a metric ([pr#18243](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: add StandbyModule and handle failed MON cluster ([pr#19744](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: don't crash when encountering an unknown PG state ([pr#18903](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: don't export metrics for dead daemon; new metrics ([pr#20506](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: fix creation of osd\_metadata metric ([pr#21530](#), Jan Fajerski)
- mgr: pybind/mgr/prometheus: fix metric type undef -> untyped ([issue#22313](#), [pr#19524](#), Ilya Margolin)

- mgr: pybind/mgr/prometheus: fix metric type undef -> untyped ([pr#18208](#), Jan Fajerski)
- mgr,pybind: pybing/mgr/prometheus: return default port if config-key get returns ... ([pr#21696](#), Jan Fajerski)
- mgr: python interface rework + enable modules to run in standby mode ([issue#21593](#), [issue#17460](#), [pr#16651](#), John Spray, Sage Weil)
- mgr: quieten logging on missing OSD stats ([pr#20485](#), John Spray)
- mgr,rbd: mgr/dashboard: added iSCSI IOPS/throughput metrics ([issue#21391](#), [pr#18653](#), Jason Dillaman)
- mgr,rbd: mgr/dashboard: fix duplicate images listed on iSCSI status page ([issue#21017](#), [pr#17055](#), Jason Dillaman)
- mgr: reconcile can\_run checks and selftest ([pr#21607](#), John Spray, Kefu Chai)
- mgr: remove a few junk lines ([pr#20005](#), John Spray)
- mgr: remove unused static files from dashboard module ([pr#16762](#), John Spray)
- mgr: request daemon's metadata when receiving a report from an unknown server ([issue#21687](#), [pr#18484](#), Chang Liu)
- mgr,rgw: mgr/dashboard: RGW page ([pr#19512](#), Chang Liu)
- mgr,rgw: prometheus: Implement rgw\_metadata metric ([pr#21383](#), Boris Ranto)
- mgr: safety checks on pyThreadState usage ([pr#18093](#), John Spray)
- mgr: set explicit thread name ([issue#21404](#), [pr#17756](#), John Spray)
- mgr: silence warning from -Wsign-compare ([pr#17881](#), Jos Collin)
- mgr: skip first non-zero incremental in PGMap::apply\_incremental() ([issue#21773](#), [pr#18347](#), Aleksei Gutikov)
- mgr/status: output to stdout, not stderr ([issue#24175](#), [pr#22135](#), John Spray)
- mgr: store declared\_types in MgrSession ([issue#21197](#), [pr#17932](#), John Spray)
- mgr: systemd: Wait 10 seconds before restarting ceph-mgr ([issue#23083](#), [pr#20533](#), Wido den Hollander)
- mgr,tests: mgr/dashboard: skip data pool testcase for none-bluestore clusters ([pr#21004](#), Tatjana Dehler)
- mgr,tests: mgr/dashboard\_v2: fix test\_perf\_counters\_mgr\_get ([pr#20916](#), Tiago Melo)

- mgr,tests: qa: add new prometheus test to rados/mgr suite ([pr#20047](#), John Spray)
- mgr,tests: qa: configure zabbix properly before selftest ([issue#22514](#), [pr#19634](#), John Spray)
- mgr,tests: qa: fix mgr \_load\_module helper ([pr#18685](#), John Spray)
- mgr,tools: mgr/iostat: implement 'ceph iostat' as a mgr plugin ([pr#20100](#), Mohamad Gebai)
- mgr: use new style config opts + add metadata ([pr#17374](#), John Spray)
- mgr/zabbix: Fix wrong log message ([pr#21237](#), Gu Zhongyan)
- mgr/zabbix: monitoring template improvements ([pr#19901](#), Marc Schoechlin)
- mon: Add ceph osd get-require-min-compat-client command ([pr#19015](#), hansbogert)
- mon: add 'ceph osd pool get erasure allow\_ec\_overwrites' command ([pr#21102](#), Mykola Golub)
- mon: add MMonHealth back ([issue#22462](#), [pr#20528](#), Kefu Chai)
- mon: add mon\_health\_preluminous\_compat\_warning ([pr#16902](#), Sage Weil)
- mon: a few conversions to monotonic clock ([pr#18595](#), Patrick Donnelly)
- mon: align lspools output ([pr#19597](#), Jos Collin)
- mon: allow cluter and debug logs to go to stderr, with appropriate prefix ([pr#19385](#), Sage Weil)
- mon: cache reencoded osdmmaps ([issue#23713](#), [pr#21605](#), Sage Weil, Xiaoxi CHEN)
- mon: centralized config ([pr#20172](#), Sage Weil)
- mon: "ceph osd crush rule rename" support ([pr#17029](#), xie xingguo)
- mon: check monitor address configuration ([pr#18073](#), Li Wang)
- mon: clean up cluster logging on mon events ([issue#22082](#), [pr#18822](#), John Spray)
- mon: cleanups to optracker code ([pr#21371](#), John Spray)
- mon: cleanup unused option mon\_health\_data\_update\_interval ([pr#17728](#), Yao Guotao)
- mon: common/options: set max\_background\_jobs instead of max\_background\_compactions ([pr#18397](#), Kefu Chai)
- mon: Compress the warnings of pgs not scrubbed or deep-scrubbed ([pr#17295](#), Zhi Zhang)
- mon: do not use per\_pool\_sum\_delta to show recovery summary ([issue#22727](#),

pr#20009, Chang Liu)

- mon: don't blow away bootstrap-mgr on upgrades ([issue#20950](#), [pr#18399](#), John Spray)
- mon: double mon\_mgr\_mkfs\_grace from 60s -> 120s ([pr#20955](#), Sage Weil)
- mon: Drop redundant access specifier, etc (cleanup) ([pr#19028](#), Shinobu Kinjo)
- mon: dump percent\_used PGMap field as float ([pr#20439](#), John Spray)
- mon: dump servicemap along with MgrStatMonitor dump info ([pr#18760](#), Zhi Zhang)
- mon: expand cap validity check for mgr, osd, mds ([issue#22525](#), [pr#21311](#), Jing Li, Sage Weil)
- mon: final luminous compatset feature and osdmap flag ([pr#17333](#), Sage Weil)
- mon: fix commands advertised during mon cluster upgrade ([pr#16871](#), Sage Weil)
- mon: fix dropping mgr metadata for active mgr (#21260) ([issue#21260](#), [pr#17571](#), John Spray)
- mon: fix "fs new" pool metadata update, tests ([issue#20959](#), [pr#16954](#), Greg Farnum)
- mon: fix legacy health checks in 'ceph status' during upgrade; fix jewel-x upgrade combo ([pr#16967](#), Sage Weil)
- mon: fix mgr using auth\_client\_required policy ([pr#20048](#), John Spray)
- mon: fix osd out clog message ([issue#21249](#), [pr#17525](#), John Spray)
- mon: fix slow op warning on mon, improve slow op warnings ([issue#23769](#), [pr#21684](#), Sage Weil)
- mon: fix structure of 'features' command ([pr#20115](#), Sage Weil)
- mon: fix two stray legacy get\_health() callers ([pr#17269](#), Sage Weil)
- mon: fix wrong mon-num counting logic of 'ceph features' command ([pr#16887](#), xie xingguo)
- mon: handle bad snapshot removal reqs gracefully ([issue#18746](#), [pr#20835](#), Paul Emmerich)
- mon: handle monitor lag when killing mgrs ([issue#20629](#), [pr#18268](#), John Spray)
- mon: incorrect MAX AVAIL in "ceph df" ([issue#21243](#), [pr#17513](#), liuchang0812)
- mon: invalid JSON returned when querying pool parameters ([issue#23200](#), [pr#20745](#), Chang Liu)

- mon/LogMonitor: call no\_reply() on ignored log message ([pr#22104](#), Sage Weil)
- mon: mark mgr reports as no\_reply ([issue#22114](#), [pr#21057](#), Kefu Chai)
- mon: mark mon\_allow\_pool\_delete as observed ([pr#18125](#), Dan van der Ster)
- mon: mark OSD beacons and pg\_create messages as no\_reply ([issue#22114](#), [pr#20517](#), Greg Farnum)
- mon: mon/AuthMonitor: don't validate fs authorize caps with valid\_caps() ([pr#21418](#), Joao Eduardo Luis)
- mon: mon/ConfigMonitor: clean up prepare\_command() ([pr#20911](#), Gu Zhongyan)
- mon: mon/Elector: force election epoch bump on start ([issue#20949](#), [pr#16944](#), Sage Weil)
- mon: mon/Elector: remove unused member fields start\_stamp and ack\_stamp ([pr#21091](#), runsisi)
- mon: mon/LogMonitor: "log last" should return up to n entries ([pr#18759](#), Kefu Chai)
- mon: mon/MDSMonitor: fix clang build failure ([pr#20637](#), Willem Jan Withagen)
- mon: mon,mgr: make osd\_metric more popular and report slow ops to mgr ([issue#23045](#), [pr#20660](#), lvshanchun)
- mon: mon/MgrMonitor: limit mgrmap history ([issue#22257](#), [pr#19185](#), Sage Weil)
- mon: mon/MonCommands: fix copy-and-paste error ([pr#17271](#), xie xingguo)
- mon: mon,option: set default value for mon\_dns\_srv\_name ([issue#21204](#), [pr#17539](#), Kefu Chai)
- mon: mon/OSDMonitor: add location option for "crush add-bucket" command ([pr#17125](#), xie xingguo)
- mon: mon/OSDMonitor: add 'osd crush set-all-straw-buckets-to-straw2' ([pr#18460](#), Sage Weil)
- mon: mon/OSDMonitor: add plain output for "crush class ls-osd" command ([pr#17034](#), xie xingguo)
- mon: mon/OSDMonitor: add space after \_\_func\_\_ in log msg ([pr#19127](#), Kefu Chai)
- mon: mon/OSDMonitor: Better prepare\_command\_pool\_set E2BIG error message ([pr#19944](#), Brad Hubbard)
- mon: mon/OSDMonitor.cc: fix expected\_num\_objects interpret error ([issue#22530](#), [pr#19651](#), Yang Honggang)

- mon: mon/OSDMonitor.cc : set erasure-code-profile to "" when create replicated pools ([pr#19673](#), zouaiguo)
- mon: mon/OSDMonitor: check last\_scan\_epoch instead when sending creates ([issue#20785](#), [pr#17248](#), Kefu Chai)
- mon: mon/OSDMonitor: clean up cmd 'osd tree-from' ([pr#20839](#), Gu Zhongyan)
- mon: mon/OSDMonitor: do not send\_pgCreates with stale info ([issue#20785](#), [pr#17065](#), Kefu Chai)
- mon: mon/OSDMonitor: error out if setting ruleset-\* ec profile property ([pr#17848](#), Sage Weil)
- mon: mon/OSDMonitor: fix improper input/testing range of crush somke testing ([pr#17179](#), xie xingguo)
- mon: mon/OSDMonitor: fix 'osd pg temp' unable to cleanup pg-temp ([pr#16892](#), xie xingguo)
- mon: mon/OSDMonitor: implement 'osd crush ls <node>' ([pr#16920](#), Sage Weil)
- mon: mon/OSDMonitor: kill pending upmap changes too if pool is gone ([pr#20704](#), xie xingguo)
- mon: mon/OSDMonitor: logging non-active osd id when handling osd beacon ([pr#21092](#), runsisi)
- mon: mon/OSDMonitor: make 'osd crush rule rename' idempotent ([issue#21162](#), [pr#17329](#), xie xingguo)
- mon: mon/OSDMonitor: "osd pool application get" support ([issue#20976](#), [pr#16955](#), xie xingguo)
- mon: mon/OSDMonitor: txsize should be greater or eq to prune\_interval - 1 ([pr#21430](#), Kefu Chai)
- mon: mon/PGMap: drop DISK LOG column ([pr#17617](#), Sage Weil)
- mon: mon/PGMap: fix "0 stuck requests are blocked > 4096 sec" warn ([pr#17099](#), xie xingguo)
- mon: mon/PGMap: nice numbers for 'data' section of 'ceph df' command ([pr#17368](#), xie xingguo)
- mon: mon/PGMap: Remove unnecessary header ([pr#18343](#), Shinobu Kinjo)
- mon: mon/PGMap: reweight::by\_utilization - skip DNE osds ([issue#20970](#), [pr#17064](#), xie xingguo)
- mon: mon/pgmap: update pool nearfull display ([pr#17043](#), huanwen ren)

- mon: more aggressively convert crush rulesets -> distinct rules ([pr#17508](#), John Spray, Sage Weil)
- mon: more constness ([pr#17748](#), Kefu Chai)
- mon: node ls improvement ([pr#20820](#), Gu Zhongyan)
- mon: 'node ls' mgr support ([pr#20711](#), Gu Zhongyan)
- mon: NULL check of logger before use ([pr#18788](#), Amit Kumar)
- mon,osd: dump "compression\_algorithms" in "mon metadata" ([issue#24135](#), [issue#22420](#), [pr#22004](#), Kefu Chai, Casey Bodley)
- mon: osd feature checks with 0 up osds ([issue#21471](#), [issue#20751](#), [pr#17831](#), Brad Hubbard, Sage Weil)
- mon: osdmap prune ([pr#19331](#), Joao Eduardo Luis)
- mon/OSDMonitor: cleanup: move bufferlist before use ([pr#18258](#), Shinobu Kinjo)
- mon/OSDMonitor: use new style options ([pr#18209](#), Kefu Chai)
- mon: osd/OSDMap.h: toss osd out if it has no more pending states ([pr#19642](#), xie xingguo)
- mon: paxos cleanup ([pr#20078](#), huanwen ren)
- mon/PGMap: let pg\_string\_state() return boost::optional<> ([issue#21609](#), [pr#18218](#), Kefu Chai)
- mon/PGMap: use new-style options and cleanup ([pr#18647](#), Kefu Chai)
- mon: post-luminous cleanup (part 3 of ?) ([pr#17607](#), Sage Weil)
- mon: rate limit on health check update logging ([issue#20888](#), [pr#16942](#), John Spray)
- mon: reenable timer to send digest when paxos is temporarily inactive ([issue#22142](#), [pr#19404](#), Jan Fajerski)
- mon: remove health service ([pr#20119](#), Chang Liu)
- mon: remove\_is\_write\_ready() ([pr#19191](#), Kefu Chai)
- mon: remove pre-luminous compat cruft (2 of many) ([pr#17322](#), Sage Weil)
- mon: remove unused waiting\_for\_commit ([pr#18617](#), Kefu Chai)
- mon: return directly after health\_events\_cleanup ([pr#16964](#), wang yang)
- mon: revert mds metadata argument name change ([issue#22527](#), [pr#19926](#), Patrick Donnelly)

- mon: show feature flags when printing MonSession ([pr#17535](#), Paul Emmerich)
- mon: some cleanup ([pr#17067](#), huanwen ren)
- mon,tests: vstart: set osd\_pool\_default\_erasure\_code\_profile in initial ceph.conf ([pr#21008](#), Mykola Golub)
- mon: update get\_store\_prefixes implementations ([issue#21534](#), [pr#17940](#), John Spray, huanwen ren)
- mon: update PaxosService::cached\_first\_committed in PaxosService::maybe\_trim() ([issue#11332](#), [pr#19397](#), Xuehan Xu, yupeng chen)
- mon: use ceph\_clock\_now if message is self-generated ([pr#17311](#), huangjun)
- mon: warn about using osd new instead of osd create ([issue#21023](#), [pr#17242](#), Neha Ojha)
- msg/async/AsyncConnection: remove legacy feature case handle ([pr#18469](#), Haomai Wang)
- msg/async: avoid referencing the temporary string ([pr#20640](#), Kefu Chai)
- msg/async: batch handle msg iovlen ([pr#18415](#), Jianpeng Ma)
- msg/async/dpdk: remove xsky copyright and LGPL copying ([pr#21121](#), Kefu Chai)
- msg/async/EventKqueue: assert on OOM ([pr#21488](#), Kefu Chai)
- msg/async: fix ms\_dpdk\_coremask and ms\_dpdk\_coremask conflict ([pr#18678](#), chunmei)
- msg/async:fix the incoming parameter type of EventCenter::process\_events() ([pr#20607](#), shangfufei)
- msg/async misc cleanup ([pr#18531](#), Jianpeng Ma)
- msg/async: misc cleanup ([pr#18575](#), Jianpeng Ma)
- msg/async/rdma: a tiny typo fix ([pr#18660](#), Yan Lei)
- msg/async/rdma: fix a coredump introduced by PR #18053 ([pr#18204](#), Yan Lei)
- msg/async/rdma: fix a potential coredump when handling tx\_buffers under heavy RDMA ([pr#18036](#), Yan Lei)
- msg/async/rdma: fixes crash for multi rados client within one process ([pr#16981](#), Alex Mikheev, Haomai Wang, Adir Lev)
- msg/async/rdma: fix Tx buffer leakage that can introduce “heartbeat no reply” ([pr#18053](#), Yan Lei)
- msg/async/rdma: refactor rx buffer pool allocator ([pr#17018](#), Alex Mikheev)

- msg/async/rdma: unnecessary reinitialization of an iterator ([pr#18190](#), JustL)
- msg/async: size of EventCenter::file\_events should be greater than fd ([issue#23253](#), [pr#20764](#), Yupeng Chen)
- msg/async: use bitset<> to do the popcnt ([pr#18681](#), Kefu Chai)
- msg/async: use device before checking ([pr#19738](#), Xiaoyan Li)
- msg: drop duplicate include ([pr#19623](#), /bin/bash)
- msg: drop the unnecessary polling\_stop() ([pr#17079](#), Jos Collin)
- msg: Initialize lkey,bound,port\_cnt,num\_chunk,gid\_idx ([pr#17797](#), Amit Kumar)
- msg: Initializing class members in module msg ([pr#17568](#), Amit Kumar)
- msg: reimplement sigpipe blocking ([pr#18105](#), Greg Farnum)
- msg: remove the ),it's redundant ([pr#17544](#), linxuhua)
- msg: resurrect support for !CEPH\_FEATURE\_MSG\_AUTH ([pr#19044](#), Ilya Dryomov)
- msgr: Optimization for connection establishment ([pr#16006](#), shangfufei)
- msg/simple: pass a char for reading from shutdown\_rd\_fd ([pr#19094](#), Kefu Chai)
- NVMDDevice: fix issued caused by #17002 ([pr#17112](#), Ziye Yang)
- objclass-sdk: expose \_\_cls\_init() to the world ([pr#21581](#), Kefu Chai)
- objecter: minor cleanups ([pr#19994](#), runsisi)
- os/bluestore/bluestore\_tool: Move redundant code into one method ([pr#19160](#), Shinobu Kinjo)
- os/bluestore: implement BlueRocksEnv::AreFilesSame() ([issue#21842](#), [pr#18392](#), Kefu Chai)
- os/bluestore: simplify and fix SharedBlob::put() ([issue#24211](#), [pr#22170](#), Sage Weil)
- osd: additional protection for out-of-bounds EC reads ([issue#21629](#), [pr#18088](#), Jason Dillaman)
- osd: add multiple objecter finishers ([pr#16521](#), Myoungwon Oh)
- osd: add num\_object\_manifest ([pr#20690](#), Myoungwon Oh)
- osd: add numpg\_removing metric ([pr#18450](#), Sage Weil)
- osd: add processed\_subop\_count for cls\_cxx\_subop\_version() ([issue#21964](#), [pr#18610](#), Casey Bodley)

- osd: add scrub week day constraint ([pr#18368](#), kungf)
- osd: adjust osd\_min\_pg\_log\_entries ([issue#21026](#), [pr#17075](#), J. Eric Ivancich)
- osd: allow FULL\_TRY after failsafe ([pr#17177](#), Pan Liu)
- osd: allow PG recovery scheduling preemption ([pr#17839](#), Sage Weil)
- osd: async recovery ([pr#19811](#), Neha Ojha)
- osd: avoid encoding the same log entries repeatedly for different peers ([pr#20201](#), Jianpeng Ma)
- osd: avoid the config's get\_val() overhead on the read path ([pr#20217](#), Radoslaw Zarzynski)
- osd: avoid unnecessary ref-counting across PrimaryLogPG::get\_rw\_locks ([pr#21028](#), Radoslaw Zarzynski)
- osd: be more precise about our asserts and cases when rebuilding missing sets ([issue#20985](#), [pr#17000](#), Greg Farnum)
- osd: bring in dmclock library changes ([pr#16755](#), J. Eric Ivancich)
- osd: bring in latest dmclock library updates ([pr#17997](#), J. Eric Ivancich)
- osd: cap snaptrimq\_len at 2^32 ([pr#21107](#), Kefu Chai)
- osd: change log level when withholding pg creation ([issue#22440](#), [pr#20167](#), Dan van der Ster)
- osd: change op delayed state to 'waiting for scrub' ([pr#19295](#), kungf)
- osd: Change shard digests to hex like object info digests ([pr#21362](#), David Zafman)
- osd: change the conditional in \_update\_calc\_stats ([pr#13383](#), Zhiqiang Wang)
- osd: check feature bits when encoding objectstore\_perf\_stat\_t ([pr#20378](#), Kefu Chai)
- osd: clean up dup index logic; maintain index flag logic in fewer places ([pr#16829](#), J. Eric Ivancich)
- osd: clean up pre-luminous compat cruft (part 1 of many) ([pr#17247](#), Sage Weil)
- osd: cleanups ([pr#17753](#), Kefu Chai)
- osdc/Objecter: using coarse\_mono instead ([pr#18473](#), Haomai Wang)
- osdc/Objecter: use std::shared\_mutex instead of boost::shared\_mutex ([issue#23910](#), [pr#21702](#), Abhishek Lekshmanan)

- osd: correct several spell errors in comments ([pr#21064](#), songweibin)
- osdc: Remove a bit too redundant public label ([pr#19466](#), Shinobu Kinjo)
- osdc: self-managed snapshot helper should catch decode exception ([issue#24103](#), [issue#24000](#), [pr#21958](#), Jason Dillaman)
- osd: debug dispatch\_context cases where queries not sent ([pr#20917](#), Sage Weil)
- osd: Deleting dead code PrimaryLogPG.cc ([pr#17339](#), Amit Kumar)
- osd: don't crash on empty snapset ([issue#23851](#), [pr#21058](#), Mykola Golub, Igor Fedotov)
- osd: Don't include same header twice ([pr#18319](#), Shinobu Kinjo)
- osd: Don't initialize pointers by NULL or 0 ([pr#18311](#), Shinobu Kinjo)
- osd: don't memcpy hobject\_t in PrimaryLogPG::close\_op\_ctx() ([pr#21029](#), Radoslaw Zarzynski)
- osd: don't process ostream strings when not debugging ([pr#20298](#), Mark Nelson)
- osd: drop redundant comment ([pr#20347](#), songweibin)
- osd: Drop the unused code in OSD::\_collect\_metadata ([pr#17131](#), Luo Kexue)
- osd: drop unused osd\_disk\_tp related options ([pr#21339](#), Gu Zhongyan)
- osd: eliminate ineffective container operations ([pr#19099](#), Igor Fedotov)
- osd: enumerate device names in a simple way ([pr#18453](#), Sage Weil)
- osd: exit(1) directly without lock if init fails ([pr#16647](#), Kefu Chai)
- osd: fast dispatch of peering events and pg\_map + osd sharded wq refactor ([pr#19973](#), Sage Weil)
- osd: fine-grained statistics of logical object space usage ([pr#15199](#), xie xingguo)
- osd: Fix assert when checking missing version ([issue#21218](#), [pr#20410](#), David Zafman)
- osd: fix a valgrind issue (conditional jump depends on uninitialized value) ([issue#22641](#), [pr#19874](#), Myoungwon Oh)
- osd: fix bug which cause can't erase OSDShardPGSlot ([pr#21771](#), Jianpeng Ma)
- osd: fix build\_initial\_pg\_history ([issue#21203](#), [pr#17423](#), w11979, Sage Weil)
- osd: fix crash caused by divide by zero in heartbeat code ([pr#21373](#), Piotr Dałek)

- osd: fix dpdk memzon mz\_name setting issue ([pr#19809](#), chunmei Liu)
- osd: fix dpdk runtime issue based on spdk/dpdk libarary ([pr#19559](#), chunmei Liu)
- osd: fix dpdk worker references issue ([pr#19886](#), chunmei Liu)
- osd: Fixes for osd\_scrub\_during\_recovery handling ([issue#18206](#), [pr#17039](#), David Zafman)
- osd: fix out of order caused by letting old msg from down osd be processed ([issue#22570](#), [pr#19796](#), Mingxin Liu)
- osd: fix \_process handling for pg vs slot race ([pr#21745](#), Sage Weil)
- osd: fix recovery reservation bugs, and implement remote reservation preemption ([pr#18485](#), Sage Weil)
- osd: fix replica/backfill target handling of REJECT ([issue#21613](#), [pr#18070](#), Sage Weil)
- osd: fix reqid assignment for reply messages in OpRequest ([pr#17060](#), Yingxin Cheng)
- osd: fix s390x build failure ([issue#23238](#), [pr#20969](#), Nathan Cutler)
- osd: fix typos and some cleanups ([pr#19211](#), Enming Zhang)
- osd: fix unordered read bug (for chunked object) ([issue#22369](#), [pr#19464](#), Myoungwon Oh)
- osd: fix waiting\_for\_peered vs flushing ([issue#21407](#), [pr#17759](#), Sage Weil)
- osd: flush operations for chunked objects ([pr#19294](#), Myoungwon Oh)
- osd: generalize queueing and lock interface for OpWq ([pr#16030](#), Myoungwon Oh, Kefu Chai, Samuel Just)
- osd: get loadavg per cpu for scrub load threshold check ([pr#17718](#), kungf)
- osd: get rid off extent map in object\_info ([pr#19616](#), Igor Fedotov)
- osd: hold lock while accessing recovery\_needs\_sleep ([issue#21566](#), [pr#18022](#), Neha Ojha)
- osd: Improve recovery stat handling by using peer\_missing and missing\_loc info ([issue#22837](#), [pr#20220](#), Sage Weil, David Zafman)
- osd: Improve size scrub error handling and ignore system attrs in xattr checking ([issue#20243](#), [issue#18836](#), [pr#16407](#), David Zafman)
- osd: include front\_iface+back\_iface in metadata ([issue#20956](#), [pr#16941](#), John Spray)

- osd: Initialization of data members ([pr#17691](#), Amit Kumar)
- osd: Initialization of pointer cls ([pr#17115](#), amitkuma)
- osd: Initializing start\_offset,last\_offset,offset ([pr#19333](#), Amit Kumar)
- osd: initial minimal efforts to clean up PG interface ([pr#17708](#), Sage Weil)
- osd: introduce sub-chunks to erasure code plugin interface ([issue#19278](#), [pr#15193](#), Myna Vajha)
- osd: kill snapdirs ([pr#17579](#), Sage Weil)
- osd: Make dmclock's anticipation timeout be configurable ([pr#18827](#), Taewoong Kim)
- osd: make operations on ReplicatedBackend::in\_progress\_ops more effective ([pr#19035](#), Igor Fedotov)
- osd: make PG::\*Force\* event structs public ([pr#21312](#), Willem Jan Withagen)
- osd: make scrub no deadline when max interval is zero ([pr#18354](#), kungf)
- osd: make scrub right now when pg stats\_invalid is true ([pr#17884](#), kungf)
- osd: make scrub wait for ec read/modify/writes to apply ([issue#23339](#), [pr#20944](#), Sage Weil)
- osd: make snapmapper warn+clean up instead of assert ([issue#22752](#), [pr#20040](#), Sage Weil)
- osd: make stat\_bytes and stat\_bytes\_used counters PRI0\_USEFUL ([issue#21981](#), [pr#18637](#), Yao Zongyou)
- osd: make the PG's SORTBITWISE assert a more generous shutdown ([issue#20416](#), [pr#18047](#), Greg Farnum)
- osd: Making use of find to reduce computational complexity ([pr#19732](#), Shinobu Kinjo)
- osd: migrate PGLOG\_\* macros to constexpr ([issue#20811](#), [pr#19352](#), Jesse Williamson)
- osd: minor optimizations for op wq ([pr#17704](#), Sage Weil, J. Eric Ivancich)
- osd: min\_pg\_log\_entries == max == pg\_log\_dups\_tracked ([pr#20394](#), Sage Weil)
- osd: misc cleanups ([pr#17430](#), songweibin)
- osd: miscellaneous cleanups ([pr#21431](#), songweibin)
- osd: more debugging for snapmapper bug ([issue#21557](#), [pr#19366](#), Sage Weil)
- osd: object added to missing set for backfill, but is not in recovering, error!

([issue#18162](#), [pr#18145](#), David Zafman)

- osd: only exit if \*latest\* map(s) say we are destroyed ([issue#22673](#), [pr#19988](#), Sage Weil)
- osd: Only scan for omap corruption once ([issue#21328](#), [pr#17705](#), David Zafman)
- osd,os,io: Initializing C\_ProxyChunkRead members,queue,request ([pr#19336](#), amitkuma)
- osd: pass ops\_blocked\_by\_scrub() to requeue\_scrub() ([pr#20319](#), Kefu Chai)
- osd: pass pool options to ObjectStore on pg create ([issue#22419](#), [pr#19486](#), Sage Weil)
- osd/PG: fix clang build vs private state events ([pr#18217](#), Sage Weil)
- osd/PG: handle flushed event directly ([pr#19441](#), wumingqiao)
- osd/PrimaryLogPG: derr when object size becomes over osd\_max\_object\_size ([pr#19049](#), Shinobu Kinjo)
- osd: process \_scan\_snaps() with all snapshots with head ([issue#22881](#), [issue#23909](#), [pr#21546](#), David Zafman)
- osd: publish osdmap to OSDService before starting wq threads ([issue#21977](#), [pr#21623](#), Sage Weil)
- osd: pull latest dmclock subtree ([pr#19345](#), J. Eric Ivancich)
- osd: put peering events in main sharded wq ([pr#18752](#), Sage Weil)
- osd: put pg removal in op\_wq ([pr#19433](#), Sage Weil)
- osd: reduce all\_info map find to get primary ([pr#19425](#), kungf)
- osd: refcount for manifest object (redirect, chunked) ([pr#19935](#), Myoungwon Oh)
- osd: remove cost from mclock op queues; cost not handled well in dmclock ([pr#21428](#), J. Eric Ivancich)
- osd: Remove double space ([pr#19296](#), Shinobu Kinjo)
- osd: remove duplicated "commit\_queued\_for\_journal\_write" in OpTracker ([issue#23440](#), [pr#21018](#), ashitakasam)
- osd: remove duplicated function ec\_pool in pg\_pool\_t ([pr#18059](#), Chang Liu)
- osd: Remove redundant local variable declaration ([pr#19812](#), Shinobu Kinjo)
- osd: Remove unnecessary headers ([pr#19735](#), Shinobu Kinjo)
- osd: remove unused ReplicatedBackend::objects\_read\_async() ([pr#18779](#), Kefu Chai)

- osd: remove unused variable in do\_proxy\_write ([pr#17391](#), Myoungwon Oh)
- osd: replace mclock subop opclass w/ rep\_op opclass; combine duplicated code ([pr#18194](#), J. Eric Ivancich)
- osd: replace vectors\_equal() with operator==(vector<T>, vector<T>) ([pr#18064](#), Kefu Chai)
- osd: request new map from PG when needed ([issue#21428](#), [pr#17795](#), Josh Durgin)
- osd: resend osd\_pgtemp if it's not acked ([issue#23610](#), [pr#21310](#), Kefu Chai)
- osd: Revert use of dmclock message feature bit since not yet finalized ([pr#21398](#), J. Eric Ivancich)
- osd,rgw,librbd: SCA fixes ([pr#18495](#), Danny Al-Gaaf)
- osd: set min\_version to newest version in maybe\_force\_recovery ([pr#17752](#), Xinze Chi)
- osd: Sign in early SIGHUP signal ([issue#22746](#), [pr#19958](#), huanwen ren)
- osd: silence maybe-uninitialized false positives ([pr#19820](#), Yao Zongyou)
- osd: silence warnings from -Wsign-compare ([pr#17872](#), Jos Collin)
- osd: skip dumping logical devices ([pr#20740](#), songweibin)
- osd: speed up get\_key\_name ([issue#21026](#), [pr#17071](#), J. Eric Ivancich)
- osd: s/random\_shuffle()/shuffle()/ [\(pr#19872](#), Willem Jan Withagen, Kefu Chai, Greg Farnum)
- osd: subscribe osdmaps if any pending pgs ([issue#22113](#), [pr#18916](#), Kefu Chai)
- osd: subscribe to new osdmap while waiting\_for\_healthy ([issue#21121](#), [pr#17244](#), Sage Weil)
- osd: support class method whitelisting within caps ([pr#19786](#), Jason Dillaman)
- osd: treat successful and erroneous writes the same for log trimming ([issue#22050](#), [pr#20827](#), Josh Durgin)
- osd: two cleanups ([pr#20830](#), songweibin)
- osd: update dmclock library w git subtree pull ([pr#17737](#), J. Eric Ivancich)
- osd: update info only if new\_interval ([pr#17437](#), Kefu Chai)
- osd: update store with options after pg is created ([issue#22419](#), [pr#20044](#), Kefu Chai)
- osd: use dmclock library client\_info\_f function dynamically ([pr#17063](#), bspark)

- osd: use existing osd\_required variable for messenger policy ([pr#20223](#), Yan Jun)
- osd: use prefix increment for non trivial iterator ([pr#19097](#), Kefu Chai)
- osd: Use specializations, typedefs instead ([pr#19354](#), Shinobu Kinjo)
- osd: Warn about objects with too many omap entries ([pr#16332](#), Brad Hubbard)
- os/filestore/HashIndex.h: fixed a typo in comment ([pr#17685](#), yaoguotao)
- os: Initializing uninitialized members aio\_info ([pr#17066](#), amitkuma)
- os: Removing dead code from LFNIndex.cc ([pr#17297](#), Amit Kumar)
- prometheus: Handle the TIME perf counter type metrics ([pr#21749](#), Boris Ranto)
- pybind: add return note in rbd.pyx ([pr#21768](#), Zheng Yin)
- pybind/ceph\_daemon: expand the order of magnitude of ([issue#23962](#), [pr#21836](#), Guan yunfei)
- pybind: fix chart size become bigger when refresh ([issue#20746](#), [pr#16857](#), Yixing Yan)
- pybind: mgr/dashboard: fix rbd's pool sub menu ([pr#16774](#), yanyx)
- pybind,rbd: pybind/rbd: support open the image by image\_id ([pr#19361](#), songweibin)
- pybind: remove unused get\_ceph\_version() ([pr#17727](#), Kefu Chai)
- qa: add cbt repo parameter ([pr#18543](#), Neha Ojha)
- qa: Add cephmetrics suite ([pr#18451](#), Zack Cerza)
- qa: add upgrade/luminous-x suite ([pr#17160](#), Yuri Weinstein)
- qa/crontab: run the perf-basic suite every day ([pr#21252](#), Neha Ojha)
- qa: Decreased amount of jobs on master, kraken, luminous runs ([pr#17069](#), Yuri Weinstein)
- qa: install collectl with cbt task ([pr#19324](#), Neha Ojha)
- qa: mimic-dev1 backports to avoid trusty nodes ([pr#19600](#), Kefu Chai)
- qa: preserve cbt task results ([pr#19364](#), Neha Ojha)
- qa: qa/ceph-disk: enlarge the simulated SCSI disk ([issue#22136](#), [pr#19199](#), Kefu Chai)
- qa/suites/perf-basic: add desc regarding test machines ([pr#21183](#), Neha Ojha)
- qa/suites/rados/multimon/tasks/mon\_lock\_with\_skew: whitelist PG ([pr#17004](#), Sage

Weil)

- qa/suites/rados/perf: add optimized settings ([pr#17786](#), Neha Ojha)
- qa/suites/rados/perf: add workloads ([pr#18573](#), Neha Ojha)
- qa/suites/rados/verify/validator/valgrind: whitelist PG ([pr#17005](#), Sage Weil)
- qa/suites/upgrade/jewel-x/parallel: tolerate laggy mgr ([pr#17227](#), Sage Weil)
- qa/suites/upgrade/kraken-x: fixes ([pr#16881](#), Sage Weil)
- qa/suites/upgrade/luminous-x fixes ([pr#22101](#), Sage Weil)
- qa/tests - Added options to use both cases: mon.a and installer.0 ([pr#19745](#), Yuri Weinstein)
- qa/tests - Fixed typo in crontab entry ([pr#21482](#), Yuri Weinstein)
- qa/tests: fixed typo ([pr#21728](#), Yuri Weinstein)
- qa/tests - minor clean ups and made perf-suite run 3 times, so we can... ([pr#21309](#), Yuri Weinstein)
- qa/tests - one more typo fixed :( ([pr#21483](#), Yuri Weinstein)
- qa/tests: removed rest suite from the mix ([pr#21743](#), Yuri Weinstein)
- qa: wait longer for osd to flush pg stats ([issue#24321](#), [pr#22288](#), Kefu Chai)
- qa/workunits/ceph-disk: -no-mon-config ([pr#21956](#), Kefu Chai)
- rados: make ceph\_perf\_msgr\_client work for multiple jobs ([issue#22103](#), [pr#18877](#), Jeegn Chen)
- rbd: add deep cp CLI method ([pr#19996](#), songweibin)
- rbd: add group rename methods ([issue#22981](#), [pr#20577](#), songweibin)
- rbd: add notrim option to rbd map ([pr#21056](#), Hitoshi Kamei)
- rbd: add parent info when moving child into trash bin ([pr#19280](#), songweibin)
- rbd: adjusted “lock list” JSON and XML formatted output ([pr#19900](#), Jason Dillaman)
- rbd: adjusted “showmapped” JSON and XML formatted output ([pr#19937](#), Mykola Golub)
- rbd: allow remove all unprotected snapshots ([issue#23126](#), [pr#20608](#), songweibin)
- rbd: allow trash rm/purge when pool quota is full used ([pr#20697](#), songweibin)
- rbd: backport of mimic bug fixes ([issue#24009](#), [issue#24008](#), [pr#21930](#), Jason

Dillaman)

- rbd: check if an image is already mapped before rbd map ([issue#20580](#), [pr#16517](#), Jing Li)
- rbd: children list should support snapshot id optional ([issue#23399](#), [pr#20966](#), Jason Dillaman)
- rbd: cleanup handling of IEC byte units ([pr#21564](#), Jason Dillaman)
- rbd: clean up warnings when mirror commands used on non-setup pool ([issue#21319](#), [pr#17636](#), Jason Dillaman)
- rbd: cls/journal: ensure tags are properly expired ([issue#21960](#), [pr#18604](#), Jason Dillaman)
- rbd: cls/journal: fixed possible infinite loop in expire\_tags ([issue#21956](#), [pr#18592](#), Jason Dillaman)
- rbd: cls/journal: possible infinite loop within tag\_list class method ([issue#21771](#), [pr#18270](#), Jason Dillaman)
- rbd: cls/rbd: group\_image\_list incorrectly flagged as RW ([issue#23388](#), [pr#20939](#), Jason Dillaman)
- rbd: cls/rbd: metadata\_list not honoring max\_return parameter ([issue#21247](#), [pr#17499](#), Jason Dillaman)
- rbd: cls/rbd: Silence gcc7 maybe-uninitialized warning ([pr#18504](#), Brad Hubbard)
- rbd: common/options, librbd/Utils: refactor RBD feature validation ([pr#20014](#), Sage Weil)
- rbd: disk usage on empty pool no longer returns an error message ([issue#22200](#), [pr#19045](#), Jason Dillaman)
- rbd: do not show title if there is no group snapshot ([pr#20311](#), songweibin)
- rbd: don't overwrite the error code from the remove action ([pr#20481](#), Jason Dillaman)
- rbd: drop unnecessary using declaration, etc ([pr#19005](#), Shinobu Kinjo)
- rbd: eager-thick provisioning support ([pr#18317](#), Hitoshi Kamei)
- rbd: export/import image-meta when we export/import an image ([pr#17134](#), PCzhangPC)
- rbd: filter out UserSnapshotNamespace in do\_disk\_usage ([pr#20532](#), songweibin)
- rbd: fix crash during map when “rw” option is specified ([issue#21808](#), [pr#18313](#), Peter Keresztes Schmidt)

- rbd: fix logically dead code in function list\_process\_image ([pr#16971](#), Luo Kexue)
- rbd: fix rbd children listing when child is in trash ([issue#21893](#), [pr#18483](#), songweibin)
- rbd: fix thread\_offsets calculation of rbd bench ([pr#20590](#), Hitoshi Kamei)
- rbd: group misc cleanup and update rbd man page ([pr#20199](#), songweibin)
- rbd: group snapshot rename ([pr#12431](#), Victor Denisov)
- rbd: implement image qos in tokenbucket algorithm ([pr#17032](#), Dongsheng Yang)
- rbd: import with option -export-format 2 fails to protect snapshot ([issue#23038](#), [pr#20613](#), songweibin)
- rbd: improve 'import-diff' corrupt input error messages ([issue#18844](#), [pr#21249](#), Jason Dillaman)
- rbd: Initializing m\_finalize\_ctx ([pr#17563](#), Amit Kumar)
- rbd: introduce commands of "image-meta ls/rm" ([pr#16591](#), PCzhangPC)
- rbd: journal: limit number of appends sent in one librados op ([issue#23526](#), [pr#21157](#), Mykola Golub)
- rbd: journal: trivial cleanup ([pr#19317](#), Shinobu Kinjo)
- rbd: krbd: include sys/sysmacros.h for major, minor and makedev ([pr#20773](#), Ilya Dryomov)
- rbd: krbd: rewrite "already mapped" code ([pr#17638](#), Ilya Dryomov)
- rbd: librados/snap\_set\_diff: don't assert on empty snapset ([pr#20648](#), Mykola Golub)
- rbd: librbd: create+truncate for whole-object layered discards ([issue#23285](#), [pr#20809](#), Ilya Dryomov)
- rbd: librbd: make rename request complete with filtered code ([issue#23068](#), [pr#20507](#), Mykola Golub)
- rbd: librbd misc cleanup ([pr#18419](#), Jianpeng Ma)
- rbd: librbd: skip head object map update when deep copying object beyond image size ([pr#21586](#), Mykola Golub)
- rbd: librbd: sync flush should re-use existing async flush logic ([pr#18403](#), Jason Dillaman)
- rbd: librbd,test: address coverity false positives ([pr#17825](#), Amit Kumar)

- rbd: mimic: librbd: deep copy optionally support flattening cloned image ([issue#22787](#), [pr#22038](#), Mykola Golub)
- rbd: mimic: rbd-mirror: optionally support active/active replication ([pr#22105](#), Jason Dillaman)
- rbd: mimic: rbd-mirror: potential deadlock when running asok 'flush' command ([issue#24141](#), [pr#22039](#), Mykola Golub)
- rbd-mirror: additional thrasher testing ([pr#21697](#), Jason Dillaman)
- rbd-mirror: clean up spurious error log messages ([issue#21961](#), [pr#18601](#), Jason Dillaman)
- rbd-mirror: cluster watcher should ensure it has latest OSD map ([issue#22461](#), [pr#19550](#), Jason Dillaman)
- rbd-mirror: ensure unique service daemon name is utilized ([pr#19492](#), Jason Dillaman)
- rbd-mirror: fix potential infinite loop when formatting status message ([issue#22932](#), [pr#20349](#), Mykola Golub)
- rbd-mirror: forced promotion can result in incorrect status ([issue#21559](#), [pr#17979](#), Jason Dillaman)
- rbd-mirror: ImageMap memory leak fixes ([pr#19163](#), Venky Shankar)
- rbd-mirror: Improve data pool selection when creating images ([pr#18006](#), Adam Wolfe Gordon)
- rbd-mirror: integrate image map policy as incremental step to active-active ([pr#21300](#), Jason Dillaman)
- rbd-mirror: introduce basic image mapping policy ([issue#18786](#), [pr#15691](#), Venky Shankar)
- rbd-mirror: missing lock when re-sending update\_sync\_point ([pr#19011](#), Mykola Golub)
- rbd-mirror: persist image map timestamp ([pr#19338](#), Venky Shankar)
- rbd-mirror: primary image should register in remote, non-primary image's journal ([issue#21561](#), [pr#18136](#), Jason Dillaman)
- rbd-mirror: properly translate remote tag mirror uuid for local mirror ([issue#23876](#), [pr#21657](#), Jason Dillaman)
- rbd-mirror: removed dedicated thread from image deleter ([issue#15322](#), [pr#19000](#), Jason Dillaman)
- rbd-mirror: rename asok hook to match image name when not replaying ([issue#23888](#),

- pr#21682, Jason Dillaman)
- rbd-mirror: rollback state transitions in image policy (pr#19577, Venky Shankar)
  - rbd-mirror: Set the data pool correctly when creating images (issue#20567, pr#17073, Adam Wolfe Gordon)
  - rbd-mirror: simplify notifications for image assignment (issue#15764, pr#16642, Jason Dillaman)
  - rbd-mirror: strip environment/CLI overrides for remote cluster (issue#21894, pr#18490, Jason Dillaman)
  - rbd-mirror: support deferred deletions of mirrored images (pr#19536, Jason Dillaman)
  - rbd-mirror: sync image metadata when transferring remote image (issue#21535, pr#18026, Jason Dillaman)
  - rbd-mirror: track images in policy map in support of A/A (issue#18786, pr#15788, Venky Shankar)
  - rbd-mirror: update asok hook name on image rename (issue#20860, pr#16998, Mykola Golub)
  - rbd-mirror: use next transition state to check transition completeness (pr#18969, Venky Shankar)
  - rbd-nbd: allow to unmap by image or snap spec (pr#19666, Mykola Golub)
  - rbd-nbd: bug fix when running in container (issue#22012, issue#22011, pr#18663, Li Wang)
  - rbd-nbd: certain kernels may not discover resized block devices (issue#22131, pr#18947, Jason Dillaman)
  - rbd-nbd: cleanup for NBDServer shut down (pr#17283, Pan Liu)
  - rbd-nbd: fix ebusy when do map (issue#23528, pr#21142, Li Wang)
  - rbd-nbd: fix generic option issue (issue#20426, pr#17375, Pan Liu)
  - rbd-nbd: output format support for list-mapped command (pr#19704, Mykola Golub)
  - rbd-nbd: support optionally setting device timeout (issue#22333, pr#19436, Mykola Golub)
  - rbd: null check before pool\_name use (pr#18790, Amit Kumar)
  - rbd: output notifyOp request name when watching (pr#20551, shun-s)
  - rbd: parallelize “rbd ls -l” (pr#15579, Piotr Dałek)

- rbd: pool\_percent\_used should not divided by 100 ([pr#20795](#), songweibin)
- rbd: properly pass ceph global command line args to subprocess ([pr#19821](#), Mykola Golub)
- rbd: pybind/rbd: add deep\_copy method ([pr#19406](#), Mykola Golub)
- rbd: pybind/rbd: fix metadata functions error handling ([issue#22306](#), [pr#19337](#), Mykola Golub)
- rbd: python bindings fixes and improvements ([issue#23609](#), [pr#21304](#), Ricardo Dias)
- rbd: rbd-ggate: fix parsing ceph global options ([pr#19822](#), Mykola Golub)
- rbd: rbd-ggate: fix syntax error ([pr#19919](#), Willem Jan Withagen)
- rbd: rbd-ggate: make list command produce valid xml format output ([pr#19823](#), Mykola Golub)
- rbd: rbd-ggate: small fixes and improvements ([pr#19679](#), Mykola Golub)
- rbd: rbd-ggate: tool to map images on FreeBSD via GEOM Gate ([pr#15339](#), Mykola Golub)
- rbd: rbd:introduce rbd bench rw(for read and write mix) test ([pr#17461](#), PCzhangPC)
- rbd: rbd: set a default value for options in nbd map ([pr#20529](#), songweibin)
- rbd: replace positional\_path parameter with arg\_index in get\_path() ([pr#19722](#), songweibin)
- rbd: replace trash delay option, add rbd trash purge command ([pr#18323](#), Theofilos Mouratidis)
- rbd: resource agent needs to be executable ([issue#22980](#), [issue#22362](#), [pr#20397](#), Tim Bishop)
- rbd:rm unnecessary conversion from string to char\* in image-meta function ([pr#17184](#), PCzhangPC)
- rbd: show read:write proportion in the infomation of readwrite bench test ([pr#18249](#), PCzhangPC)
- rbd: snap limit should't be set smaller than the number of existing snaps ([pr#16597](#), PCzhangPC)
- rbd: support cloning an image from a non-primary snapshot ([issue#18480](#), [pr#19724](#), Jason Dillaman)
- rbd: support iterating over metadata items when listing ([issue#21179](#), [pr#17532](#), Jason Dillaman)

- rbd: support lock\_timeout in rbd mapping ([pr#21344](#), Dongsheng Yang)
- rbd: support osd\_request\_timeout in rbd map command ([issue#23073](#), [pr#20792](#), Dongsheng Yang)
- rbd: switched from legacy to new-style configuration options ([issue#20737](#), [pr#16737](#), Jason Dillaman)
- rbd,tests: qa: additional krbd discard test cases ([pr#20499](#), Ilya Dryomov)
- rbd,tests: qa: fix POOL\_APP\_NOT\_ENABLED warning in krbd:unmap suite ([pr#16966](#), Ilya Dryomov)
- rbd,tests: qa: introduce rbd-mirror thrasher to existing tests ([issue#18753](#), [pr#21541](#), Jason Dillaman)
- rbd,tests: qa: krbd\_exclusive\_option.sh: add lock\_timeout test case ([pr#21522](#), Ilya Dryomov)
- rbd,tests: qa: krbd\_fallocate.sh: add notrim test case ([pr#21513](#), Ilya Dryomov)
- rbd,tests: qa: krbd huge-image test ([pr#20692](#), Ilya Dryomov)
- rbd,tests: qa: krbd latest-osdmap-on-map test ([pr#20591](#), Ilya Dryomov)
- rbd,tests: qa: krbd msgr-segments test ([pr#20714](#), Ilya Dryomov)
- rbd,tests: qa: krbd parent-overlap test ([pr#20721](#), Ilya Dryomov)
- rbd,tests: qa: krbd whole-object-discard test ([pr#20750](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: add krbd BLKROSET test ([pr#18652](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: enable generic/050 and generic/448 ([pr#18795](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: enable xfstests blockdev tests ([pr#17621](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: exclude shared/298 ([pr#17971](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: rbd\_xfstests job overhaul ([pr#17346](#), Ilya Dryomov)
- rbd,tests: qa/suites/rbd: fewer socket failures ([pr#19617](#), Sage Weil)
- rbd,tests: qa/suites/rbd: miscellaneous test fixes ([issue#21251](#), [pr#17504](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: segregated v1 image format tests ([issue#22738](#), [pr#20729](#), Jason Dillaman)
- rbd,tests: qa/suites/rbd: set qemu task time\_wait param ([pr#21131](#), Mykola Golub)

- rbd,tests: qa/tasks/cram: include /usr/sbin in the PATH for all commands ([pr#18793](#), Ilya Dryomov)
- rbd,tests: qa/tasks/rbd: run all xfstests runs to completion ([pr#18583](#), Ilya Dryomov)
- rbd,tests: qa/workunits/rbd: fix cli\_generic test\_purge for rbd default format 1 ([pr#20389](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: fixed variable name for resync image id ([issue#21663](#), [pr#18097](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: fix issues within permissions test ([issue#23043](#), [pr#20491](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: pool create may fail for small cluster ([pr#18067](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: potential race in mirror disconnect test ([issue#23938](#), [pr#21733](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: relax greps to support upgrade formatting change ([issue#21181](#), [pr#17559](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: remove sanity check in journal.sh test ([pr#20490](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: remove sanity check in test\_admin\_socket.sh ([pr#21116](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: remove “trash purge -threshold” test ([issue#22803](#), [pr#20170](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: simplify split-brain test to avoid potential race ([issue#22485](#), [pr#19604](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: switch devstack tempest to 17.2.0 tag ([issue#22961](#), [pr#20599](#), Jason Dillaman)

- rbd,tests: qa/workunits/rbd: switch devstack to pike release ([pr#20124](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: test data pool is mirrored correctly ([pr#17062](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: unnecessary sleep after failed remove ([pr#18619](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: use command line option to specify watcher asok ([issue#20954](#), [pr#16917](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: wait for demote status is propagated ([pr#19073](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: wait for status propagated only if daemon started ([pr#19082](#), Mykola Golub)
- rbd,tests: rbd/test: add snap protection test for ex/import ([pr#20689](#), songweibin)
- rbd,tests: stop.sh: use -no-mon-config when trying to unmap rbd devices ([pr#21020](#), Mykola Golub)
- rbd,tests: test: address coverity false positives ([pr#17803](#), Amit Kumar)
- rbd,tests: test/cls\_rbd: mask newer feature bits to support upgrade tests ([issue#21217](#), [pr#17509](#), Jason Dillaman)
- rbd,tests: test/librados\_test\_stub: always create copy of buffers passed to operation ([pr#21074](#), Mykola Golub)
- rbd,tests: test/librbd: added update\_features RPC message to test\_notify ([issue#21936](#), [pr#18561](#), Jason Dillaman)
- rbd,tests: test/librbd: clean up for several mock function tests ([pr#18952](#), Jason Dillaman)
- rbd,tests: test/librbd: Do not instantiate TrimRequest template class ([pr#19402](#), Boris Ranto)
- rbd,tests: test/librbd: ensure OutOfOrder test has enough concurrent management ops ([pr#21436](#), Mykola Golub)
- rbd,tests: test/librbd: fix mock method macro of set\_journal\_policy ([pr#17216](#), Yan Jun)
- rbd,tests: test/librbd: fix race condition with OSD map refresh ([issue#20918](#), [pr#16877](#), Jason Dillaman)
- rbd,tests: test/librbd: fix valgrind memory leak warning ([pr#17187](#), Mykola Golub)

- rbd,tests: test/librbd: initialize on\_finish,locker,force,snap\_id ([pr#17800](#), Amit Kumar)
- rbd,tests: test/librbd: make fsx build on non-linux platform ([pr#16939](#), Mykola Golub)
- rbd,tests: test/librbd: memory leak in recently added test ([pr#18478](#), Mykola Golub)
- rbd,tests: test/librbd: rbd-ggate mode for fsx ([pr#19315](#), Mykola Golub)
- rbd,tests: test/librbd: test metadata\_set/remove is applied ([pr#18288](#), Mykola Golub)
- rbd,tests: test/librbd: TestMirroringWatcher unit tests should ignore duplicates ([issue#21029](#), [pr#17078](#), Jason Dillaman)
- rbd,tests: test/librbd: utilize unique pool for cache tier testing ([issue#11502](#), [pr#20486](#), Jason Dillaman)
- rbd,tests: test/librbd: valgrind warning in TestMockManagedLockBreakRequest.DeadLockOwner ([pr#18940](#), Mykola Golub)
- rbd,tests: test/pybind/rbd: skip test\_deep\_copy\_clone if layering not enabled ([pr#20295](#), Mykola Golub)
- rbd,tests: test/rbd: cli\_generic fails if v1 image format or deep-flatten disabled ([issue#22950](#), [pr#20364](#), songweibin)
- rbd,tests: test/rbd\_mirror: fix valgrind warnings in unittest ([pr#19016](#), Mykola Golub)
- rbd,tests: test/rbd-mirror: image map policy test ([pr#19320](#), Venky Shankar)
- rbd,tests: test/rbd-mirror: improve coverage for dead instance handling ([pr#21403](#), Jason Dillaman)
- rbd,tests: test/rbd\_mirror: “use of uninitialised value” valgrind warning ([pr#19437](#), Mykola Golub)
- rbd,tools: rbd-fuse: make sure PATH\_MAX is defined ([pr#18615](#), Roberto Oliveira)
- rbd,tools: rbd-replay: remove boost dependency ([pr#21202](#), Kefu Chai)
- rbd: tools/rbd: use steady clock in bencher ([pr#20008](#), Mohamad Gebai)
- rbd: ‘trash list -long’ will return a failure on non-cloned images ([pr#19540](#), Jason Dillaman)
- rbd: ‘trash ls -l’ will display column titles if existed non-USER trash image only ([pr#21343](#), songweibin)

- rbd: unified way to map images using different drivers ([pr#19711](#), Mykola Golub)
- rbd: use different logic to disturb thread's offset in bench seq test ([pr#17218](#), PCzhangPC)
- Revert "ceph-fuse: Delete inode's bufferhead was in Tx state would le... ([pr#21976](#), "Yan, Zheng")
- Revert "msg/async/rdma: fix multi cephcontext confllicting" ([pr#16980](#), Haomai Wang)
- Revert "os/bluestore: compensate for bad freelistmanager size/blocks metadata" ([pr#17275](#), Xie Xingguo)
- rgw: ability to list bucket contents in unsorted order for efficiency ([pr#21026](#), J. Eric Ivancich)
- rgw: abort multipart if upload meta object doesn't exist ([pr#19918](#), fang yuxiang)
- rgw: Access RGWConf through RGWEnv ([pr#17432](#), Jos Collin)
- rgw: add "Accept-Ranges" to response header of Swift API ([issue#21554](#), [pr#17967](#), Tone Zhang)
- rgw: add a default redirect field for zones ([pr#9571](#), Yehuda Sadeh)
- rgw: add an option to clear all usage entries ([pr#19322](#), Abhishek Lekshmanan)
- rgw: add an option to recalculate user stats ([issue#23335](#), [pr#20853](#), Abhishek Lekshmanan)
- rgw: add buffering filter to compression for fetch\_remote\_obj ([issue#23547](#), [pr#21479](#), Casey Bodley)
- rgw: add cors header rule check in cors option request ([issue#22002](#), [pr#18556](#), yuliyang)
- rgw: Add dynamic resharding documentation ([issue#21553](#), [pr#15941](#), Orit Wasserman)
- rgw: add logs if get\_data returns error in RGWPutObj::execute ([pr#18642](#), Zhang Shaowen)
- rgw: add metadata and data sync related cmd into radosgw-admin usage ([pr#18921](#), lvshanchun)
- rgw: add missing override in list\_keys\_init() ([pr#17254](#), Jos Collin)
- rgw: add radosgw-admin sync error trim to trim sync error log ([pr#19854](#), fang yuxiang)
- rgw: add reshard commands ([issue#21617](#), [pr#18180](#), Orit Wasserman)

- rgw: address warnings due to incorrect format code ([pr#18796](#), J. Eric Ivancich)  
rgw: Add retry\_raced\_bucket\_write
  - rgw: add rewrite cmd and options into radosgw-admin usage and doc ([pr#18918](#), Enming Zhang)
  - rgw: add ssl support to beast frontend ([issue#22832](#), [pr#20464](#), Casey Bodley)
  - rgw: add support for Swift's per storage policy statistics ([issue#17932](#), [pr#12704](#), Radoslaw Zarzynski)
  - rgw: add support for Swift's reversed account listings ([issue#21148](#), [pr#17320](#), Radoslaw Zarzynski)
  - rgw: add support for tagging and other conditionals in policy ([pr#17094](#), Abhishek Lekshmanan)
  - rgw: add tail tag to track tail instance ([issue#20234](#), [pr#16145](#), Yehuda Sadeh)
  - rgw: add tenant to shard\_id in RGWDeleteLC::execute() ([pr#10460](#), Wei Qiaomiao)
  - rgw: add time skew check in function parse\_v4\_auth\_header ([issue#22418](#), [pr#19476](#), Bingyin Zhang)
- rgw: Add try\_refresh\_bucket\_info function
- rgw: add xml output header in RGWCopyObj\_ObjStore\_S3 response msg ([issue#22416](#), [pr#19475](#), Enming Zhang)
  - rgw: adjust log format for lifecycle ([pr#19576](#), Bingyin Zhang)
  - rgw: admin api - add ability to sync user stats from admin api ([issue#21301](#), [pr#17589](#), Nathan Johnson)
  - rgw: Admin API Support for bucket quota change ([issue#21811](#), [pr#18324](#), Jeegn Chen)
  - rgw: admin rest api shouldn't return error when getting user's stats if the user hasn't create any bucket ([pr#21551](#), Zhang Shaowen)
  - rgw: allow beast frontend to listen on specific IP address ([issue#22778](#), [pr#20000](#), Yuan Zhou)
  - rgw: Allow swift acls to be deleted ([issue#22897](#), [pr#20471](#), Marcus Watts)
  - rgw: avoid logging keystone revocation messages when not configured ([issue#21400](#), [pr#17775](#), Abhishek Lekshmanan)
  - rgw: aws4 auth supports PutBucketRequestPayment ([issue#23803](#), [pr#21569](#), Casey Bodley)

- rgw: AWS v4 authorization work when INIT\_MULTIPART is chunked ([issue#22129](#), [pr#18956](#), Jeegn Chen)
- rgw: beast frontend can listen on multiple endpoints ([issue#22779](#), [pr#20188](#), Casey Bodley)
- rgw: beast frontend no longer experimental ([pr#21272](#), Casey Bodley)
- rgw: Better ERANGE error message ([issue#22351](#), [pr#20023](#), Brad Hubbard)
- rgw: break sending data-log list infinitely ([issue#20951](#), [pr#16926](#), fang.yuxiang)
- rgw: bucket resharding should not update bucket ACL or user stats ([issue#22742](#), [issue#22124](#), [pr#20038](#), Orit Wasserman)
- rgw: Cache on the barrelhead ([issue#22517](#), [pr#19581](#), Adam C. Emerson)
- rgw: Cache Register! ([issue#22604](#), [issue#22603](#), [pr#20144](#), Adam C. Emerson)
- rgw: can't download object with range when compression enabled ([issue#22852](#), [pr#20226](#), fang yuxiang)
- rgw: ceph-dencoder: add missing begin\_iter & end\_iter item for RGWObjManifest ([pr#19509](#), wangsongbo)
- rgw: ceph-dencoder: add support for cls\_rgw\_lc\_obj\_head ([pr#18920](#), Yao Zongyou)
- rgw: ceph-dencoder: add support for RGWLifecycleConfiguration ([pr#18959](#), wangsongbo)
- rgw: change ObjectCache::lru from deque back to list ([issue#22560](#), [pr#19768](#), Casey Bodley)
- rgw: changes to support ragweed ([pr#13644](#), Yehuda Sadeh)
- rgw: Check bucket CORS operations in policy ([issue#21578](#), [pr#18000](#), Adam C. Emerson)
- rgw: Check bucket GetBucketLocation in policy ([issue#21582](#), [pr#18002](#), Adam C. Emerson)
- rgw: Check bucket Website operations in policy ([issue#21597](#), [pr#18024](#), Adam C. Emerson)
- rgw: check going\_down() when lifecycle processing ([issue#22099](#), [pr#18846](#), Yao Zongyou)
- rgw: Check payment operations in policy ([issue#21389](#), [pr#17742](#), Adam C. Emerson)
- rgw: check read\_op.read return value in RGWRados::copy\_obj\_data ([pr#18962](#), Enming Zhang)

- rgw: civetweb fixes for v1.1 upgrade ([pr#21123](#), Abhishek Lekshmanan)
- rgw: clean code with helper function dump\_header\_if\_nonempty ([pr#18979](#), Xinying Song)
- rgw: clean up and fix some bugs for encryption ([issue#21581](#), [pr#17882](#), Enming Zhang)
- rgw: cleanup MIN macro with std::min ([pr#17546](#), Jiaying Ren)
- rgw: cleanup unused parameters in RGWRados::copy\_obj\_data ([pr#18917](#), Enming Zhang)
- rgw: cloud sync fixes ([pr#21648](#), Yehuda Sadeh)
- rgw: cls/log: cls\_log\_list always returns next marker ([issue#20906](#), [pr#17024](#), Casey Bodley)
- rgw: cls/rgw: fix bi\_log\_iterate\_entries return wrong truncated ([issue#22737](#), [pr#20021](#), Tianshan Qu)
- rgw: cls/rgw: Initialization of uninitialized members ([pr#16932](#), amitkuma)
- rgw: cls/rgw: mtime in rgw\_bucket\_dir\_entry\_meta not really decoded ([issue#22148](#), [pr#18981](#), Yao Zongyou)
- rgw: cls/rgw: remove unused variable bl ([pr#19570](#), Yao Zongyou)
- rgw: cls/rgw: trim all usage entries in cls\_rgw ([issue#22234](#), [pr#19131](#), Abhishek Lekshmanan)
- rgw: cls\_rgw: use more effective container operations in get\_obj\_vals ([pr#19272](#), Xinying Song)
- rgw: comparison between signed and unsigned integer expressions ([pr#21105](#), ashitakasam)
- rgw: consolidate code that implements hashing algorithms ([pr#18248](#), J. Eric Ivancich)
- rgw: copy object add response error messages ([pr#18291](#), Enming Zhang)
- rgw: correct comment in function parse\_credentials ([pr#19275](#), Bingyin Zhang)
- rgw: correct log output for metadata section name in RGWListBucketIndexesCR ([pr#19508](#), Xinying Song)
- rgw: Correct permission evaluation to allow only admin users to work with Roles ([pr#20332](#), Pritha Srivastava)
- rgw: correct typo refity to refit ([pr#19064](#), Bingyin Zhang)

- rgw: correct typo UNKOWN to UNKNOWN ([pr#19273](#), Bingyin Zhang)
- rgw: create sync-module instance when execute radosgw-admin data sync run ([issue#22080](#), [pr#18898](#), lvshanchun)
- rgw: create sync-module instance when radosgw-admin sync run ([pr#20611](#), lvshanchun)
- rgw: curl\* reuse and for debian, use openssl not gnutls ([pr#20635](#), Marcus Watts)
- rgw: Data encryption is not follow the AWS agreement ([pr#15994](#), hechuang)
- rgw: datalog list support -shard-id and -marker ([pr#20649](#), Tianshan Qu)
- rgw: data sync: set num\_shards when building full maps ([issue#22083](#), [pr#18852](#), Abhishek Lekshmanan)
- rgw: Delete to\_string functions. stringify defined in include/stringify.h can provide the same feature ([pr#18522](#), zhangwen)
- rgw: disable dynamic resharding in multisite environment ([issue#21725](#), [pr#18184](#), Orit Wasserman)
- rgw: do not reflect period if not current ([issue#22844](#), [pr#20212](#), Tianshan Qu)
- rgw: do not update all gateway caches upon creation of system obj w/ exclusive flag ([pr#19384](#), J. Eric Ivancich)
- rgw: don't change rados object's mtime when update olh ([issue#21743](#), [pr#18214](#), Shasha Lu)
- rgw: don't hold data\_lock over frontend io ([pr#20621](#), Casey Bodley)
- rgw: don't leak S3 LDAPHelper ([pr#12427](#), Matt Benjamin)
- rgw: dont log EBUSY errors in 'sync error list' ([issue#22473](#), [pr#19580](#), Casey Bodley)
- rgw: dont reuse stale RGWObjectCtx for get\_bucket\_info() ([issue#21506](#), [pr#17916](#), Casey Bodley)
- rgw: don't write bucket\_header when it is not changed in bucket\_link/unlink ([pr#17356](#), Shasha Lu)
- rgw: don't write bucket\_header when it is not changed in rgw\_bucket\_prepare\_op ([pr#18763](#), Xinying Song)
- rgw: download object might fail for local invariable uninitialized ([issue#23146](#), [pr#20612](#), fang yuxiang)
- rgw: drop a repeated statement for encode\_xml() ([pr#20195](#), luomuyao)

- rgw: drop commented functions ([pr#19671](#), Jos Collin)
- rgw: drop dump\_uri\_from\_state() which isn't used anymore ([pr#19924](#), Radoslaw Zarzynski)
- rgw: drop iter in rgw\_op.cc ([pr#19583](#), Bingyin Zhang)
- rgw: drop marker in RGWL::process() ([pr#19591](#), Bingyin Zhang)
- rgw: drop outdated function doc ([pr#18370](#), Jiaying Ren)
- rgw: drop "realm remove" in radosgw-admin ([pr#18212](#), Shasha Lu)
- rgw: drop redundant RGW\_OP\_STAT\_OBJ check ([pr#19933](#), Bingyin Zhang)
- rgw: drop the unnecessary handling of Swift's X-Storage-Policy on objects ([pr#16383](#), Jiaying Ren)
- rgw: drop the unused function init\_anon\_user() ([pr#16874](#), Radoslaw Zarzynski)
- rgw: Drop unnecessary return ([pr#17520](#), Jos Collin)
- rgw: drop unused function apply\_epoch ([pr#17593](#), Shasha Lu)
- rgw: drop unused iter in XMLObj::find\_first ([pr#19709](#), luomuyao)
- rgw: drop unused variable bucket\_instance\_ids ([pr#19708](#), Bingyin Zhang)
- rgw: drop unused variable in copy\_obj\_data() ([pr#18477](#), Enming Zhang)
- rgw: drop unused vector elements ([pr#19815](#), Bingyin Zhang)
- rgw: drop useless includes in rgw\_{main.cc, common.h} ([pr#19109](#), Jiaying Ren)
- rgw: drop useless lines ([pr#19817](#), Bingyin Zhang)
- rgw: drop useless type conversion ([pr#19824](#), Bingyin Zhang)
- rgw: drop variable bl in rgw\_op.cc ([pr#19584](#), Bingyin Zhang)
- rgw: Drop #warning TODO ([issue#19851](#), [pr#17012](#), Jos Collin)
- rgw: dump Last-Modified in Swift's responses for GET/HEAD on container ([issue#20883](#), [pr#16757](#), Radoslaw Zarzynski)
- rgw: enable 'qlen' & 'qactive' performance counters ([pr#20842](#), Mark Kogan)
- rgw: encoding fixes ([issue#23779](#), [pr#21500](#), Yehuda Sadeh)
- rgw: Error check on return of read\_line() ([pr#17880](#), Amit Kumar)
- rgw: es module: set compression type correctly ([issue#22758](#), [pr#20796](#), Abhishek Lekshmanan)

- rgw: evaluate the correct bucket action for GetACL bucket operation ([issue#21013](#), [pr#17050](#), Abhishek Lekshmanan)
- rgw: exit early if rgw\_bucket\_set\_attrs() fails ([pr#17041](#), dengxiafubi)  
rgw: Expire entries in bucket info cache
- rgw\_file: fix write error when the write offset overlaps ([issue#21455](#), [pr#17809](#), Yao Zongyou)
- rgw: fix a bug in rgw cache in delete\_system\_obj and get\_system\_obj ([pr#10992](#), zhangshaowen)
- rgw: fix accessing expired memory in PrefixableSignatureHelper ([issue#21085](#), [pr#17206](#), Radoslaw Zarzynski)
- rgw: fix a typo in comment ([pr#19608](#), luomuyao)
- rgw: fix a typo in comment ([pr#20164](#), luomuyao)
- rgw: fix a typo in comment ([pr#20355](#), luomuyao)
- rgw: fix a typo in rgw\_perms[] ([pr#20024](#), luomuyao)
- rgw: fix bilog entries on multipart complete ([issue#21772](#), [pr#18271](#), Casey Bodley)
- rgw: fix BZ 1500904, stale bucket index entry remains after obj delete ([pr#18709](#), J. Eric Ivancich)
- rgw: fix chained cache invalidation to prevent cache size growth ([issue#22410](#), [pr#19455](#), Mark Kogan)
- rgw: Fix closing tag for Prefix ([pr#17663](#), Shasha Lu)
- rgw: fix cls\_bucket\_head result order consistency ([pr#18700](#), Tianshan Qu)
- rgw: fix collect()'s return in coroutine ([pr#19606](#), Xinying Song)
- rgw: fix command argument error for radosgw-admin ([issue#21723](#), [pr#18175](#), Yao Zongyou)
- rgw: fix 'copy part' without 'x-amz-copy-source-range' ([issue#22729](#), [pr#20002](#), Malcolm Lee)
- rgw: fix 'copy part' without 'x-amz-copy-source-range' when compression enabled ([issue#23196](#), [pr#20686](#), fang yuxiang)
- rgw: fix crash with rgw\_run\_sync\_thread false ([issue#20448](#), [pr#20769](#), Orit Wasserman)
- rgw: Fix dereference of empty optional ([issue#21962](#), [pr#18602](#), Adam C. Emerson)

- rgw: fix error handling for GET with ?torrent ([issue#23506](#), [pr#21576](#), Casey Bodley)
- rgw: fix error handling in Browser Uploads ([pr#15054](#), Radoslaw Zarzynski)
- rgw: fix error handling in ListBucketIndexesCR ([issue#21735](#), [pr#18198](#), Casey Bodley)
- rgw: fixes for multisite replication of encrypted objects ([issue#20668](#), [issue#20671](#), [pr#16612](#), Casey Bodley)
- rgw: fix extra\_data\_len handling in PutObj filters ([issue#21895](#), [pr#18489](#), Casey Bodley)
- rgw: fix for empty query string in beast frontend ([issue#22797](#), [pr#20120](#), Casey Bodley)
- rgw: fix for issue #21647 ([issue#23859](#), [pr#21647](#), Yehuda Sadeh)
- rgw: fix for pause in beast frontend ([issue#21831](#), [pr#18402](#), Casey Bodley)
- rgw: fix for usage truncated flag ([pr#20926](#), Yehuda Sadeh, Greg Farnum, Robin H. Johnson)
- rgw: Fix getter function names in RGWEnv ([pr#18377](#), Jos Collin)
- rgw: fix GET website response error code ([issue#22272](#), [pr#19236](#), Dmitry Plyakin)
- rgw: fix handling of ENOENT in RGWRadosGetOmapKeysCR ([pr#19878](#), Casey Bodley)
- rgw: fix index cancel op miss update header ([pr#20396](#), Tianshan Qu)
- rgw: Fix infinite call for bi list when resharding a bucket ([issue#22721](#), [pr#21584](#), Orit Wasserman)
- rgw: fix lc process only schedule the first item of lc objects ([issue#21022](#), [pr#17061](#), Shasha Lu)
- rgw:fix list objects with marker wrong result when bucket is enable versioning ([issue#21500](#), [pr#17934](#), yuliyang)
- rgw: fix memory fragmentation problem reading data from client ([pr#20724](#), Marcus Watts)
- rgw: Fix multisite Synchronization failed when read and write delete ... ([issue#22804](#), [pr#20814](#), Niu Pengju)
- rgw: fix not responding when receiving SIGHUP signal ([pr#16854](#), Yao Zongyou)
- rgw: fix null pointer crush ([pr#18861](#), Sibei Gao)
- rgw: fix obj copied from remote gateway acl full\_control issue ([issue#20658](#),

pr#16127, Enming Zhang)

- rgw: fix opslog cannot record remote\_addr ([issue#20931](#), [pr#16860](#), Jiaying Ren)
- rgw: fix opslog can't record referrer when using curl as client ([issue#20935](#), [pr#16863](#), Jiaying Ren)
- rgw: fix opslog uri as per Amazon s3 ([issue#20971](#), [pr#16958](#), Jiaying Ren)
- rgw: fix radosgw-admin bucket rm with --purge-objects and --bypass-gc ([issue#22122](#), [issue#19959](#), [pr#18922](#), Aleksei Gutikov)
- rgw: fix radosgw-admin quota enable return value bug ([issue#21608](#), [pr#18057](#), baixueyu)
- rgw: fix radosgw linkage with WITH\_RADOSGW\_BEAST\_FRONTEND=OFF ([issue#23680](#), [pr#21380](#), Casey Bodley)
- rgw: fix recursive lock ([pr#19430](#), Tianshan Qu)
- rgw: fix resource leak in rgw\_bucket.cc and rgw\_user.cc ([issue#21214](#), [pr#17353](#), Luo Kexue)
- rgw: fix return value of auth v2/v4 ([issue#22439](#), [pr#19310](#), Bingyin Zhang)
- rgw: fix rewrite a versioning object create a new object bug ([issue#21984](#), [pr#18662](#), Enming Zhang)
- rgw: fix rewrite options usage text ([pr#18968](#), Jos Collin)
- rgw: fix RGWCompletionManager get\_next stuck after going down ([issue#22799](#), [pr#20095](#), Tianshan Qu)
- rgw: fix RGWLibIO did not init RGWEnv ([pr#19065](#), Tianshan Qu)
- rgw: fix s3 website redirection error ([pr#19252](#), yuliyang)
- rgw: fix s3website redirect location string length ([pr#19826](#), yuliyang)
- rgw: fix Swift container naming rules ([issue#19264](#), [pr#13992](#), Robin H. Johnson)
- rgw: Fix swift object expiry not deleting objects ([issue#22084](#), [pr#18821](#), Pavan Rallabhandi)
- rgw: fix sync status conflict with cloud sync ([pr#21425](#), Casey Bodley)
- rgw: fix the bug of radowgw-admin zonegroup set requires realm ([issue#21583](#), [pr#19061](#), lvshanchun)
- rgw: fix the max-uploads parameter not work ([issue#22825](#), [pr#20158](#), Xin Liao)
- rgw: fix the return type is wrong ([pr#19773](#), hechuang)

- rgw: fix total\_time to msec as per AWS S3 ([pr#17541](#), Jiaying Ren)
- rgw: fix typo anynoymous to anonymous ([pr#19281](#), Bingyin Zhang)
- rgw: fix typo compete to complete ([pr#19675](#), Bingyin Zhang)
- rgw: Fix typo in comment ([pr#21032](#), Simran Singhal)
- rgw: fix typo in GetOmapKeysCR ([pr#19713](#), lvshanchun)
- rgw: fix typo signle to single ([pr#19517](#), Bingyin Zhang)
- rgw: fix typo woild to would ([pr#19472](#), Bingyin Zhang)
- rgw: Fix use after free in IAM policy parser ([pr#16823](#), Adam C. Emerson)
- rgw: fix use of libcurl with empty header values ([issue#23663](#), [pr#21358](#), Casey Bodley)
- rgw: format logs in file rgw\_lc.cc ([pr#19615](#), Bingyin Zhang)
- rgw: format rgw\_bucket\_dir\_header in ceph-dencoder ([pr#19753](#), Bingyin Zhang)
- rgw: gc use aio ([pr#20546](#), Yehuda Sadeh)

rgw: Handle stale bucket info in RGWDeleteBucketPolicy  
rgw: Handle stale bucket info in RGWDeleteBucketWebsite  
rgw: Handle stale bucket info in RGWPutBucketPolicy  
rgw: Handle stale bucket info in RGWPutMetadataBucket  
rgw: Handle stale bucket info in RGWSetBucketVersioning  
rgw: Handle stale bucket info in RGWSetBucketWebsite

- rgw: honor the tenant part of rgw\_bucket during comparisons ([issue#20897](#), [pr#16796](#), Radoslaw Zarzynski)
- rgw: iam policy printing cleanups ([pr#18961](#), Kefu Chai)
- rgw: Ignoring the returned error ([pr#17907](#), Amit Kumar)
- rgw: implement ipv4 aws:SourceIp condition for bucket policy ([pr#19167](#), yuliyang)
- rgw: improve handling of Swift's error messages and limits ([issue#17938](#), [issue#21169](#), [issue#17935](#), [issue#17934](#), [issue#17936](#), [pr#15369](#), Radoslaw Zarzynski)
- rgw: improve sync status: display behind bucket shards ([pr#20027](#), lvshanchun)
- rgw: improve sync status ([pr#19573](#), lvshanchun)
- rgw: include SSE-KMS headers in encrypted upload response ([issue#21576](#), [pr#17999](#), Casey Bodley)
- rgw: incorporate the Transfer-Encoding fix for CivetWeb ([issue#21015](#), [pr#17072](#), Radoslaw Zarzynski)

- rgw: Initialization of epoch,len ([pr#17722](#), Amit Kumar)
- rgw: Initialize is\_master, max\_aio, size ([pr#16933](#), amitkuma)
- rgw: Initializes uninitialized members ([pr#16855](#), Amit Kumar)
- rgw: init oldest period after setting run\_sync\_thread ([issue#21996](#), [pr#18664](#), Orit Wasserman, Casey Bodley)
- rgw: keep compression type consistent between parts of s3 Multipart ([pr#19740](#), fang yuxiang)
- rgw: keystone: bump up logging when error is received ([issue#22151](#), [pr#18985](#), Abhishek Lekshmanan)
- rgw:lc fix expiration time ([issue#21533](#), [pr#17824](#), Shasha Lu)
- rgw: lc support Content-MD5 request header and fix a rgw crash bug ([issue#21980](#), [pr#18534](#), Enming Zhang)
- rgw: lease\_cr->go\_down is called twice, remove the needless one ([pr#19394](#), Zhang Shaowen)
- rgw: librgw: export multitenancy support ([pr#19358](#), Tao Chen)
- rgw: librgw: fix shutdown err with resources uncleaned ([issue#22296](#), [pr#19279](#), Tao Chen)
- rgw: lifecycle omap entry was removed in abnormal situation ([pr#19921](#), fang yuxiang)
- rgw: list\_objects() honors end\_marker regardless of namespace ([issue#18977](#), [pr#15273](#), Radoslaw Zarzynski)
- rgw: loadgen fix generate random object name rgw crash issue ([issue#22006](#), [pr#18536](#), Enming Zhang)
- rgw: log the right http status code in civetweb frontend's access log ([issue#22538](#), [pr#19678](#), Yao Zongyou)
- rgw: log unlink\_instance mtime as object's mtime ([issue#18885](#), [pr#20016](#), Yehuda Sadeh)
- rgw: lttng: Trace rgw data transfer, bi entry and object header update processes ([pr#20556](#), Yang Honggang)
- rgw: make init env methods return an error ([issue#23039](#), [pr#20488](#), Abhishek Lekshmanan)
- rgw: make radosgw object stat RGW\_ATTR\_COMPRESSION dump readable ([pr#19846](#), fang yuxiang)

- rgw: mfa support ([pr#19283](#), Yehuda Sadeh)
- rgw: mimic: rgw: policy: modify s3>ListBucketMultiPartUploads to s3>ListBucketMul ([issue#24062](#), [pr#21916](#), xiangxiang)
- rgw: modify s3 type subuser access permissions fail through admin rest api ([issue#21983](#), [pr#18641](#), yuliyang)
- rgw: move all pool creation into rgw\_init\_ioctx ([issue#23480](#), [pr#21534](#), Casey Bodley)
- rgw: mrgw.sh uses instance name 'client.rgw' ([pr#18404](#), Casey Bodley)
- rgw: multisite log tracing ([pr#16492](#), Yehuda Sadeh, Casey Bodley)
- rgw,nfs: Add hint to use -o sync when mounting ([pr#16210](#), Adam Kupczyk)
- rgw: no need to deal with md5 header in get\_data ([pr#19144](#), Zhang Shaowen)
- rgw: optimize function abort\_bucket\_multiparts ([pr#19710](#), Bingyin Zhang)
- rgw: optimize function bucket\_lc\_prepare ([pr#19613](#), Bingyin Zhang)
- rgw: optimize function parse\_raw\_oid ([pr#19814](#), Bingyin Zhang)
- rgw: optimize function RGWHandler::do\_init\_permissions ([pr#19700](#), Bingyin Zhang)
- rgw: optimize memory usage in function rgw\_bucket::get\_key ([pr#19391](#), Bingyin Zhang)
- rgw: optimize next start time for lifecycle ([pr#19596](#), Bingyin Zhang)
- rgw: optimize the rgw-attr del code logic ([pr#18895](#), wangsongbo)
- rgw: optimize time skew check ([pr#19511](#), Bingyin Zhang)
- rgw: parse old rgw\_obj with namespace correctly ([issue#22982](#), [pr#20425](#), Yehuda Sadeh)
- rgw: policy: support for s3 conditionals in ListBucket Op ([pr#16628](#), Abhishek Lekshmanan)
- rgw: Potential fix for possible 500 on POST ([pr#18954](#), Adam C. Emerson)
- rgw: Prevent overflow of cached stats values ([issue#20934](#), [pr#17116](#), Aleksei Gutikov)
- rgw: proper error message when tier\_type does not exist ([issue#22469](#), [pr#19575](#), lvshanchun, Chang Liu)
- rgw: pull up beast submodule and update frontend ([pr#17923](#), Casey Bodley)
- rgw: put bucket policy panics RGW process ([issue#22541](#), [pr#19687](#), Bingyin Zhang)

- rgw: radosgw-admin abort early for user stats for empty uids ([issue#23322](#), [pr#20846](#), Abhishek Lekshmanan)
- rgw: radosgw-admin should not use metadata cache for readonly commands ([issue#23468](#), [pr#21129](#), Orit Wasserman)
- rgw: radosgw-admin zonegroup get and zone get return defaults when there is no realm ([issue#21615](#), [pr#18667](#), lvshanchun)
- rgw: radosgw: fix awsv4 header line sort order ([issue#21607](#), [pr#18046](#), Marcus Watts)
- rgw: radosgw: usage: fix bytes\_sent bug ([issue#19870](#), [pr#16834](#), Marcus Watts)
- rgw: raise log level on coroutine shutdown errors ([issue#23974](#), [pr#21791](#), Casey Bodley)
- rgw: Reinstating error codes mapping for Roles ([pr#20309](#), Pritha Srivastava)
- rgw: reject encrypted object COPY before supported ([issue#23232](#), [pr#20739](#), Jeegn Chen)
- rgw: release cls lock if taken in RGWCompleteMultipart ([issue#21596](#), [pr#18104](#), Matt Benjamin)
- rgw: Remove assertions in IAM Policy ([pr#18225](#), Adam C. Emerson)
- rgw: remove get\_system\_obj\_attrs in function RGWDeleteLC::execute and RGWDeleteCORS::execute ([pr#19582](#), Bingyin Zhang)
- rgw: remove placement\_rule from rgw\_link\_bucket() ([issue#21990](#), [pr#18657](#), Casey Bodley)
- rgw: remove redundant parenthesis in logs ([pr#19375](#), Bingyin Zhang)
- rgw: remove redundant S3AnonymousEngine ([pr#19474](#), Bingyin Zhang)
- rgw: remove redundant signature compare in LocalEngine::authenticate ([pr#19676](#), Bingyin Zhang)
- rgw: Remove the useless output when list zones ([pr#17434](#), iliul)
- rgw: remove unused cls\_user\_add\_bucket ([pr#19917](#), Yao Zongyou)
- rgw: remove unused disable\_signal\_fd ([pr#18875](#), Yao Zongyou)
- rgw: remove unused function get\_system\_obj\_attrs ([pr#19852](#), Yao Zongyou)
- rgw: Remove unused Parameter in Function RGWConf::init() ([pr#17129](#), Wen Zhang)
- rgw: remove unused param in AWSGeneralAbstractor::get\_auth\_data\_v4 ([pr#19250](#), Bingyin Zhang)

- rgw: remove unused param in get\_bucket\_instance\_policy\_from\_attr ([pr#19129](#), Bingyin Zhang)
- rgw: remove unused variables ([pr#16649](#), Zhang Lei)
- rgw: remove useless lines in RGWDeleteBucket::execute ([pr#19699](#), Bingyin Zhang)
- rgw: reshards cancel command should clear bucket resharding flag ([issue#21619](#), [pr#21120](#), Orit Wasserman)
- rgw: reshards should not update stats when linking new bucket instance ([issue#22124](#), [pr#19253](#), Orit Wasserman)
- rgw: retry CORS put/delete operations on ECANCELLED ([issue#22517](#), [pr#19601](#), Adam C. Emerson)
- rgw: return 'Access-Control-Allow-Origin' header when the set and delete bucket website through XMLHttpRequest ([pr#17632](#), yuliyang)
- rgw: return 'Access-Control-Allow-Origin' header when the set bucket versioning through XMLHttpRequest ([pr#17631](#), yuliyang)
- rgw: return bucket's location no matter which zonegroup it located in ([issue#21125](#), [pr#17250](#), Shasha Lu)
- rgw: return EINVAL if max\_keys can not convert correctly ([issue#23586](#), [pr#21285](#), yuliyang)
- rgw: Return Error if Bucket Policy Contains Undefined Action ([pr#17433](#), zhangwen)
- rgw: Returning when dst\_ioctlx.operate() returns error ([pr#17873](#), Amit Kumar)
- rgw: return valid Location element, CompleteMultipartUpload ([pr#19902](#), Matt Benjamin)
- rgw: revert PR #8765 ([pr#16807](#), fang.yuxiang)
- rgw: Revert "radosgw: fix awsv4 header line sort order." ([issue#21832](#), [pr#18381](#), Casey Bodley)
- rgw: Revert "rgw\_file: disable FLAG\_EXACT\_MATCH enforcement" ([issue#22827](#), [pr#20171](#), Matt Benjamin)
- rgw: Revert "rgw: reshards should not update stats when linking new bucket instance" ([pr#20052](#), Orit Wasserman)
- rgw: rework json/xml escape usage follow #19806 ([pr#19845](#), fang yuxiang)
- rgw: rgw-admin: check input parameters for friendly prompt ([pr#17343](#), Yao Zongyou)
- rgw: rgw-admin: check the data extra pool supports omap ([pr#18978](#), Yao Zongyou)

- rgw: rgw-admin: properly filtering bucket stats by user\_id or bucket\_name ([pr#19401](#), Yao Zongyou)
- rgw: rgw-admin: require -yes-i-really-mean-it when using -inconsistent\_index ([issue#20777](#), [pr#17185](#), Orit Wasserman)
- rgw: rgw-admin: support for processing all gc objects including unexpired ([pr#17482](#), Yao Zongyou)
- rgw: RGW: change function parameters from value to refrence ([pr#18355](#), Sibei Gao)
- rgw: RGWCivetWeb::read\_data: fix arguments to mg\_read() call ([issue#23596](#), [pr#21291](#), Nathan Cutler)
- rgw: rgw clean-up: remove unreferenced pure virtual class StreamObjData ([pr#18799](#), J. Eric Ivancich)
- rgw: rgw clean-up: remove unused var & func in RGWRados::SystemObject ([pr#18987](#), J. Eric Ivancich)
- rgw: rgw cleanup: some unnecessary function called and repeated assignment ([pr#18817](#), Enming Zhang)
- rgw: rgw cloud sync ([issue#21802](#), [pr#18932](#), lvshanchun, Yehuda Sadeh, Chang Liu, Abhishek Lekshmanan)
- rgw: RGWEnv::set() takes std::string ([issue#22101](#), [pr#18866](#), Casey Bodley)
- rgw: rgw\_file: alternate fix deadlock on lru eviction ([pr#20034](#), Matt Benjamin)
- rgw: rgw\_file: avoid evaluating nullptr for readdir offset ([pr#20145](#), Matt Benjamin)
- rgw: rgw\_file: conditionally unlink handles when direct deleted ([issue#23299](#), [pr#20834](#), Matt Benjamin)
- rgw: rgw\_file: explicit NFSv3 open() emulation ([pr#18365](#), Matt Benjamin)
- rgw: rgw\_file: fix LRU lane lock in evict\_block() ([issue#21141](#), [pr#17267](#), Matt Benjamin)
- rgw: rgw\_file: implement variant offset readdir processing ([pr#18335](#), Matt Benjamin)
- rgw: rgw\_file: introduce new fsid and rgw\_mount ([pr#15330](#), Gui Hecheng)
- rgw: rgw\_file: set s->obj\_size from bytes\_written ([issue#21940](#), [pr#18571](#), Matt Benjamin)
- rgw: rgw\_file: Silence unused-function warnings ([pr#19278](#), Brad Hubbard)
- rgw: RGW: fix a bug about inconsistent unit of comparison ([issue#21590](#), [pr#17958](#),

- gaosibei)
- rgw: rgw.iam: change '1' to '1ULL' in function print\_actions ([pr#18900](#), Bingyin Zhang)
  - rgw: rgw\_lc: add support for optional filter argument and make ID optional ([issue#19587](#), [issue#20872](#), [pr#16818](#), Abhishek Lekshmanan)
  - rgw: rgw\_lc: support for AWSv4 authentication ([pr#16734](#), Abhishek Lekshmanan)
  - rgw: rgw\_log, rgw\_file: account for new required envvars ([issue#21942](#), [pr#18572](#), Matt Benjamin)
  - rgw: Rgw master fix plus ([issue#21000](#), [issue#21003](#), [issue#20501](#), [pr#17040](#), Zhang Shaowen, Marcus Watts)
  - rgw: rgw, mon: normalize delete/remove in admin console (cleanup 22444) ([issue#14363](#), [issue#22444](#), [pr#19439](#), Jesse Williamson)
  - rgw: RGW: Multipart upload may double the quota ([issue#21586](#), [pr#17959](#), Sibei Gao)
  - rgw: rgw\_multisite: automated trimming for bucket index logs ([issue#18229](#), [pr#17761](#), Casey Bodley)
  - rgw: RGW NFS: mount cmdline example missing -osync ([pr#15855](#), Matt Benjamin)
  - rgw: RGW-NFS: Use rados cluster\_stat to report filesystem usage ([issue#22202](#), [pr#20093](#), Supriti Singh)
  - rgw: rgw\_op: Drop the Old LifecycleConfiguration from logs ([pr#16821](#), Abhishek Lekshmanan)
  - rgw: rgw\_op: exit early if object has no attrs in GetObjectTagging ([issue#21010](#), [pr#17048](#), Abhishek Lekshmanan)
  - rgw: RGWPutLC return ERR\_MALFORMED\_XML when missing <Rule> tag in lifecycle.xml ([issue#21377](#), [pr#17683](#), Shasha Lu)
  - rgw: rgw\_put\_system\_obj takes bufferlist ([pr#19897](#), Casey Bodley)
  - rgw: rgw\_rados: set\_attrs now sets the same time for BI & object ([issue#21200](#), [pr#17400](#), Abhishek Lekshmanan)
  - rgw: rgw/rgw\_op.cc: Fix error message in rgw\_user\_get\_all\_buckets\_stats ([pr#18781](#), iliul)
  - rgw: rgw: source data in 'default.rgw.buckets.data' may not be deleted after inter-bucket copy ([issue#21819](#), [pr#18369](#), baixueyu)
  - rgw: RGW: support for tagging in lifecycle policies ([pr#17305](#), Abhishek Lekshmanan)

- rgw: RGW: update S3 POST policy handling of Content-Type ([issue#20201](#), [pr#18658](#), Matt Benjamin)
- rgw: rgw: use camelcase format in request headers ([pr#19210](#), lvshanchun, Chang Liu)
- rgw: RGWUser::init no longer overwrites user\_id ([issue#21685](#), [pr#18137](#), Casey Bodley)
- rgw: S3 Bucket Policy Conditions IpAddress and NotIpAddress do not work ([issue#20991](#), [pr#17010](#), John Gibson)
- rgw: s3website error handler uses original object name ([issue#23201](#), [pr#20693](#), Casey Bodley)
- rgw: send x-amz-version-id header when upload files ([pr#18935](#), Xinying Song)
- rgw: set bucket versioning donot change versioning status if missing status in xml ([issue#21364](#), [pr#17662](#), Shasha Lu)
- rgw: set num\_shards on 'radosgw-admin data sync init' ([issue#22083](#), [pr#18883](#), Casey Bodley)
- rgw: set priority on perf counters ([pr#20006](#), John Spray)
- rgw: set sync\_from\_all as true when no value is seen ([issue#22062](#), [pr#18926](#), Abhishek Lekshmanan)
- rgw: setup locks for libopenssl ([issue#22951](#), [issue#23203](#), [pr#20390](#), Abhishek Lekshmanan, Jesse Williamson)
- rgw: share time skew check between v2 and v4 auth ([pr#20013](#), Casey Bodley)
- rgw: Silence maybe-uninitialized false positives ([pr#19274](#), Brad Hubbard)
- rgw: silence not allow register storage class specifier warning ([pr#19859](#), Yao Zongyou)
- rgw: simplify use of map::emplace in iam ([pr#18706](#), Casey Bodley)
- rgw: Small refactor and two bug fixes ([issue#21901](#), [issue#21896](#), [pr#18606](#), Adam C. Emerson)
- rgw: some cleanup for sync status ([pr#20894](#), Enming Zhang)
- rgw: stop/join TokenCache revoke thread only if started ([issue#21666](#), [pr#18106](#), Karol Mroz)
- rgw: stream metadata full sync init ([issue#18079](#), [pr#12429](#), Yehuda Sadeh)
- rgw: submodule: update Beast to ceph/ceph-master branch ([pr#19182](#), Casey Bodley)

- rgw: switch beast frontend back to stackful coroutine ([issue#20048](#), [pr#20449](#), Casey Bodley)
- rgw: sync tracing fixes ([issue#22833](#), [pr#20191](#), Yehuda Sadeh)
- rgw: tenant fixes for dynamic resharding ([issue#22046](#), [pr#18811](#), Orit Wasserman)
- rgw,tests: fix s3atests that are failing for sometime ([pr#20678](#), Vasu Kulkarni)
- rgw,tests: qa: fix overrides for openssl\_keys task ([pr#20981](#), Casey Bodley)
- rgw,tests: qa: re enable LC tests ([pr#17020](#), Abhishek Lekshmanan)
- rgw,tests: qa/rgw: add beast frontend to some rgw suites ([pr#17977](#), Casey Bodley)
- rgw,tests: qa/rgw: combine swift, s3tests, ragweed into single verify task ([pr#20756](#), Casey Bodley)
- rgw,tests: qa/rgw: disable log trim in multisite suite ([pr#19438](#), Casey Bodley)
- rgw,tests: qa/rgw: hadoop-s3a suite targets centos\_latest ([pr#17777](#), Casey Bodley)
- rgw,tests: qa/rgw: ignore errors from 'pool application enable' ([issue#21715](#), [pr#18193](#), Casey Bodley)
- rgw,tests: qa/rgw: remove some civetweb overrides for beast testing ([issue#23002](#), [pr#20440](#), Casey Bodley)
- rgw,tests: qa/rgw: renamed ssl task to openssl\_keys ([pr#20863](#), Ricardo Dias)
- rgw,tests: qa/rgw: use 'ceph osd pool application enable' on created pools ([pr#17162](#), Casey Bodley)
- rgw,tests: qa/rgw: verify suite tests civetweb with ssl ([pr#20444](#), Casey Bodley)
- rgw,tests: qa/smoke: add rgw crypto config for s3tests ([pr#17700](#), Casey Bodley)
- rgw,tests: qa/tasks/swift: add support for the "force-branch" configurable ([pr#21027](#), Radoslaw Zarzynski)
- rgw,tests: rgw, qa: integrate Tempest to verify RadosGW's compliance with Swift API ([pr#16344](#), Radoslaw Zarzynski)
- rgw,tests: test/rgw: fix test\_encrypted\_object\_sync for 3+ zones ([pr#17377](#), Casey Bodley)
- rgw: the metavariables in frontends-related config won't be expanded ([pr#19689](#), root)
- rgw,tools: tools/rgw: add script to inspect admin socket "cr dump" ([pr#15554](#), Casey Bodley)

- rgw: Torrents are not supported for objects encrypted using SSE-C ([issue#21720](#), [pr#17956](#), Zhang Shaowen)
- rgw: trim all spaces inside a metadata value ([issue#23301](#), [pr#20841](#), Orit Wasserman)
- rgw: update radosgw-admin usage with bi purge ([pr#18245](#), Yao Zongyou)
- rgw: unlink deleted bucket from bucket's owner ([issue#22248](#), [pr#20017](#), Casey Bodley)
- rgw: unreachable return in RGWRados::trim\_bi\_log\_entries ([pr#17367](#), Amit Kumar)
- rgw: update life cycle related log level ([pr#18845](#), Yao Zongyou)
- rgw: update outdated debug func name ([pr#17440](#), Jiaying Ren)
- rgw: update quota is inconsistent at add/del object with compression ([issue#22568](#), [pr#19772](#), fang yuxiang)
- rgw: update the usage read iterator in truncated scenario ([issue#21196](#), [pr#17939](#), Mark Kogan)
- rgw: update usage() with status ([pr#18178](#), Jos Collin)
- rgw: update vstart.sh to support rgw ssl port notation : '-rgw\_port 443s' ([issue#21151](#), [pr#17989](#), Mark Kogan)
- rgw: update the max-buckets when the quota is uploaded ([pr#20063](#), zhaokun)
- rgw: URL-decode S3 and Swift object-copy URLs ([issue#22121](#), [pr#19936](#), Matt Benjamin)
- rgw: url\_encode key name and instance in es sync module ([pr#20707](#), Chang Liu)
- rgw: use explicit index pool placement ([issue#22928](#), [pr#20352](#), Yehuda Sadeh)
- rgw: Use namespace for lc\_pool and roles\_pool ([issue#20177](#), [pr#16889](#), Orit Wasserman)
- rgw: Various cleanups and options update in rgw\_admin.cc ([pr#18302](#), Jos Collin)
- rgw: vstart.sh: fix mstop.sh can not stop rgw ([pr#17438](#), Jiaying Ren)
- rgw: 'zone placement' commands validate compression type ([issue#21775](#), [pr#18273](#), Casey Bodley)
- rocksdb: sync with upstream ([issue#21603](#), [pr#18262](#), Kefu Chai)
- rpm: rm macros in comments ([issue#22250](#), [pr#17070](#), Ken Dreyer)
- script/build-integration-branch: check errors ([pr#17578](#), Sage Weil)

- script/build-integration-branch: python3 compatible and pep8 clean ([pr#18035](#), Kefu Chai)
- scripts: new backport-create-issue script ([pr#21480](#), Nathan Cutler)
- selinux: Allow ceph to execute ldconfig ([pr#21974](#), Boris Ranto)
- selinux: Allow setattr on lnk sysfs files ([pr#17891](#), Boris Ranto)
- spdk: advance to upstream dc82989d ([pr#20713](#), Nathan Cutler)
- src: fix various log messages ([pr#21112](#), Gu Zhongyan)
- src/msg/rdma: fixes failure on assert in notify() ([pr#17007](#), Alex Mikheev)
- suites/cephmetrics: Add Centos 7 ([pr#18594](#), Zack Cerza)
- test: assert check for negative returns ([pr#17296](#), Amit Kumar)
- test/fio: generate db histogram to help debug rocksdb performance ([pr#16808](#), Pan Liu, Xiaoyan Li)
- test: fix bash path in shebangs (part 2) ([pr#17955](#), Alan Somers)
- test: fix CLI unit formatting tests ([pr#22260](#), Jason Dillaman)
- test: Incorrect conversion to double ([pr#18963](#), Amit Kumar)
- test/librados: reorder ASSERT\_EQ() arguments ([pr#16625](#), Yan Jun)
- test,osd,kvstore\_tool: silence warnings and prepare test buffer in the right way ([pr#18406](#), Adam C. Emerson)
- tests: bluestore/fio: Fixed problem with all objects having the same hash ([pr#17770](#), Adam Kupczyk)
- tests: CentOS 7.4 is now the latest ([pr#17776](#), Nathan Cutler)
- tests - ceph-ansible vars additions ([issue#21822](#), [pr#18378](#), Yuri Weinstein)
- tests: ceph-disk: ignore E722 in flake8 test ([issue#22207](#), [pr#19072](#), Nathan Cutler)
- tests: ceph-disk: mock get fsid ([pr#19254](#), Kefu Chai)
- tests: ceph-disk: Remove sitepackages=True ([issue#22823](#), [pr#20151](#), Brad Hubbard)
- tests: ceph-objectstore-tool: don't destroy SnapMapper until the txn is completed ([issue#23121](#), [pr#20593](#), Kefu Chai)
- tests: Changes required for teuthology's systemd support ([pr#18380](#), Zack Cerza)
- tests: Check for empty output in test\_dump\_pgstate\_history ([pr#20838](#), Brad

Hubbard)

- tests: cleanup: drop calamari tasks ([pr#17531](#), Nathan Cutler)
- tests: cleanup: drop upgrade/jewel-x/point-to-point-x ([issue#22888](#), [pr#20245](#), Nathan Cutler)
- tests: cmake,test/mgr: restructure dashboard tests and cmake related fixes ([pr#20768](#), Kefu Chai)
- tests: common/obj\_bencher: set {min,max}\_iops if runtime < 1 sec ([pr#17182](#), Kefu Chai)
- tests: c\_read\_operations.cc: Silence tautological-compare compiler warning ([pr#19953](#), Brad Hubbard)
- tests: fix uninitialized value found by coverity scan ([pr#17895](#), J. Eric Ivancich)
- tests: Increase sleep in test\_pidfile.sh ([pr#17052](#), David Zafman)
- tests: librgw\_file: remove unused using statement ([pr#17085](#), Yao Zongyou)
- tests: mark\_unfound\_lost fix and some other minor changes ([issue#21907](#), [pr#18449](#), David Zafman)
- tests: mgr/dashboard: Allow sourcing run-backend-api-tests.sh ([pr#20874](#), Sebastian Wagner)
- tests: mgr/dashboard: create venv for running tox ([pr#21490](#), Kefu Chai)
- tests: mgr/dashboard: notification queue: fix priority tests ([pr#21147](#), Ricardo Dias)
- tests: mimic: qa: fix test on “ceph fs set cephfs allow\_new\_snaps” ([pr#21830](#), Kefu Chai)
- tests: mimic: qa/workunits/rados/test\_envlibrados\_for\_rocksdb: install g++ not g++-4.7 ([pr#22117](#), Kefu Chai)
- tests: mimic: test: Need to escape parens in log-whitelist for grep ([pr#22075](#), David Zafman)
- tests: mimic: test: wait\_for\_pg\_stats() should do another check after last 13 secon... ([pr#22199](#), David Zafman)
- tests: misc: Fix bash path in shebangs ([pr#16494](#), Alan Somers)
- tests: mstart.sh: remove bashizm in /bin/sh script ([pr#18541](#), Mykola Golub)
- tests: point-to-point-x: upgrade client.1 to -x along with cluster nodes ([issue#21499](#), [pr#17910](#), Nathan Cutler)

- tests: qa: add cbt task for performance testing ([pr#17583](#), Neha Ojha)
- tests: qa: add cosbench workloads and override teuthology default settings ([pr#21710](#), Neha Ojha)
- tests/qa: Adding \$ distro mix - rgw ([pr#22070](#), Yuri Weinstein)
- tests/qa: adding rados/.. dirs ([pr#22068](#), Yuri Weinstein)
- tests: qa: add “restful” to ceph\_mgr\_modules in ceph-ansible suite ([pr#18634](#), Kefu Chai)
- tests: qa: add simple and dirty script to find ports being used ([pr#19102](#), Joao Eduardo Luis)
- tests: qa: big: add openstack.yaml ([pr#16864](#), Nathan Cutler)
- tests: qa: clean up dnsmasq task and fix EPERM error ([pr#20680](#), Casey Bodley)
- tests: qa: create\_cache\_pool no longer runs ‘pool application enable’ ([issue#21155](#), [pr#17312](#), Casey Bodley)
- tests: qa: decrease the msg\_inject\_socket\_failures from 1/500 to 1/1000 ([issue#22093](#), [pr#19542](#), Kefu Chai)
- tests: qa: disable mon-health-to-clog in upgrade test ([pr#19233](#), Kefu Chai)
- tests: qa: disable -Werror when compiling env\_librados\_test ([pr#21433](#), Kefu Chai)
- tests: qa: do not “ceph fs get cephfs” w/o a cephfs fs ([pr#18533](#), Kefu Chai)
- tests: qa: do not wait for down/out osd for pg convergence ([pr#18808](#), Kefu Chai)
- tests/qa - enabled ceph-deploy runs on mira nodes ([pr#21253](#), Yuri Weinstein)
- tests: qa: fix pool-quota related tests ([issue#21409](#), [pr#17763](#), xie xingguo)
- tests: qa: Fix shebangs on openstack scripts ([pr#16546](#), Alan Somers)
- tests: qa: reduce “mon client hunt interval max multiple” to 2 for all clients ([pr#21658](#), Kefu Chai)
- tests: qa: reduce mon-client-hunt-interval-max-multiple to 2 ([pr#18283](#), Kefu Chai)
- tests: qa: revert “qa: use config\_path property instead of literal” ([pr#17850](#), Patrick Donnelly)
- tests: qa/run-standalone.sh: set PYTHONPATH for FreeBSD also ([pr#20646](#), Kefu Chai)
- tests: qa: s/backfill/backfilling/ ([pr#18235](#), Kefu Chai)

- tests: qa/standalone: pass options using -<option-name>=<value> ([pr#19544](#), Kefu Chai)
- tests: qa/standalone: Add trap for signals to restore the kernel core pattern ([pr#17026](#), David Zafman)
- tests: qa/standalone/ceph-helpers.sh: provide argument to dirname ([issue#23805](#), [pr#21552](#), Nathan Cutler)
- tests: qa/standalone/ceph-helpers.sh: silence ceph-disk DEPRECATION\_WARNING ([pr#19478](#), Kefu Chai)
- tests: qa/standalone: extract delete\_pool() ([pr#20634](#), Kefu Chai)
- tests: qa/standalone: misc fixes ([issue#20465](#), [issue#20921](#), [pr#16709](#), David Zafman)
- tests: qa/standalone/mon/misc.sh: Add osdmap-prune tests ([issue#23621](#), [pr#21318](#), Brad Hubbard)
- tests: qa/standalone/osd/osd-mark-down: create pool to get updated osdmap faster ([pr#18191](#), huangjun)
- tests: qa/standalone: remove osd-map-max-advance related tests ([issue#22596](#), [pr#19816](#), Kefu Chai)
- tests: qa/standalone: respect \$TEMPDIR ([pr#17747](#), Kefu Chai)
- tests: qa/standalone/scrub/osd-scrub-repair.sh: add extents flag into object\_info\_t ([issue#21618](#), [pr#18094](#), xie xingguo)
- tests: qa/standalone/scrub/osd-scrub-repair.sh: drop omap\_digest flag ([pr#18150](#), xie xingguo, Sage Weil)
- tests: qa/standalone: s/delete\_erasure\_pool/delete\_erasure\_coded\_pool/ ([pr#20667](#), Kefu Chai)
- tests: qa: stop testing deprecated "ceph osd create" ([issue#21993](#), [pr#18659](#), Kefu Chai)
- tests: qa/suites: add minimal performance suite ([pr#21104](#), Neha Ojha)
- tests: qa/suites/cephmetrics: Updates for new version ([pr#21146](#), Zack Cerza)
- tests: qa/suites: change fixed-2.yaml users to get 4 openstack disks ([pr#16873](#), Sage Weil)
- tests: qa/suites: mds.0 -> mds.a ([pr#20848](#), Sage Weil)
- tests: qa/suites/rados: Disable scrub backoff ([issue#23578](#), [pr#21295](#), Brad Hubbard)

- tests: qa/suites/rados/mgr/tasks/dashboard: add MDS\_ALL\_DOWN to whitelist ([pr#21549](#), Ricardo Dias)
- tests: qa/suites/rados/mgr/tasks/dashboard\_v2: add fail\_on\_skip = false ([pr#20925](#), Ricardo Dias)
- tests: qa/suites/rados/multimon: whitelist mgr down vs clock skew test ([pr#16996](#), Sage Weil)
- tests: qa/suites/rados/singleton: more whitelist ([pr#19225](#), Kefu Chai)
- tests: qa/suites/rados/thrash-old-clients: ms\_type=simple ([issue#23922](#), [pr#21739](#), Kefu Chai)
- tests: qa/suites/rados/upgrade/jewel-x-singleton: tolerate sloppy past\_intervals ([pr#17293](#), Kefu Chai)
- tests: qa/suites/rest/basic/tasks/rest\_test: more whitelisting ([issue#21425](#), [pr#17794](#), huangjun)
- tests: qa/suites/rest/basic/tasks/rest\_test: whiltelist OSD\_DOWN ([issue#21425](#), [pr#18144](#), huangjun)
- tests: qa/suites/upgarde/jewel-x/parallel: tolerate mgr warning ([pr#17203](#), Sage Weil)
- tests: qa/suites/upgarde/jewel-x/point-to-point-x: disable app warnings ([pr#16947](#), Sage Weil)
- tests: qa/suites: whitelist SLOW\_OPS ([issue#23495](#), [pr#21324](#), Kefu Chai)
- tests: qa/tasks: Add default timeout for wait for pg clean task ([pr#21313](#), Vasu Kulkarni)
- tests: qa/tasks/ceph\_deploy: gatherkeys before mgr deploy ([pr#17224](#), Sage Weil)
- tests: qa/tasks/ceph\_manager: use set\_config on revived osd ([pr#20901](#), Neha Ojha)
- tests: qa/tasks/mgr/dashboard: Fix login expires too soon ([pr#21021](#), Sebastian Wagner)
- tests: qa/tasks: prolong revive\_osd() timeout to 6 min ([issue#21474](#), [pr#17902](#), Kefu Chai)
- tests: qa/tasks: prolong revive\_osd() timeout to 6 min ([issue#21474](#), [pr#19024](#), Kefu Chai)
- tests: qa/tasks: run cosbench using the CBT task ([pr#21656](#), Neha Ojha)
- tests: qa/tasks: update ceph-deploy task to use newer ceph-volume syntax ([pr#19244](#), Vasu Kulkarni)

- tests: qa/tests: Add additional required ceph-ansible vars due to upstream changes ([pr#17757](#), Vasu Kulkarni)
- tests: qa/tests: add ceph-deploy upgrade tests ([issue#20950](#), [pr#16826](#), Vasu Kulkarni)
- tests: qa/tests: add openstack volume info + lvs for ceph-volume ([pr#20243](#), Vasu Kulkarni)
- tests: qa/tests: Fix get\_system\_type failure due to invalid remote name ([pr#17650](#), Vasu Kulkarni)
- tests: qa/tests: fix rbd pool creation for systemd tests ([pr#17536](#), Vasu Kulkarni)
- tests: [qa/tests]: misc ceph-ansible fixes and update ([pr#17096](#), Vasu Kulkarni)
- tests: qa/tests/rados: Remove unsupported 2-size-1-min-size config ([pr#17576](#), Vasu Kulkarni)
- tests: qa/tests: use ceph-deploy stable branch for single node tests ([pr#20979](#), Vasu Kulkarni)
- tests: qa/tests: Various whitelists for smoke suite ([issue#21376](#), [pr#17680](#), Vasu Kulkarni)
- tests: qa/tests: Wip ceph deploy upgrade ([pr#17651](#), Vasu Kulkarni)
- tests: qa/workunits/rados/test\_large\_omap\_detection: Scrub pgs instead of OSDs ([pr#21410](#), Brad Hubbard)
- tests: qa/workunits: silence py warnings for ceph-disk tests ([issue#22154](#), [pr#19075](#), Kefu Chai)
- tests: rados: Copy payload in ceph\_perf\_msgr\_client ([issue#22100](#), [pr#18862](#), Jeegn Chen)
- tests: rados: Intializing members class StriperTest ([pr#16843](#), amitkuma)
- tests: remove TestPGLog ASSERT\_DEATH test ([issue#23504](#), [pr#21117](#), Nathan Cutler)
- tests: run-standalone.sh improve error message ([pr#17093](#), David Zafman)
- tests: run-standalone.sh skip core\_pattern if already set ([pr#17098](#), David Zafman)
- tests: test/admin\_socket\_output: add -vstart=path/to/asok option ([pr#20371](#), Kefu Chai)
- tests: test/admin\_socket\_output: better error reporting ([pr#20409](#), Brad Hubbard)
- tests: test/admin\_socket\_output: switch to std::experimental::filesystem

([pr#20307](#), Kefu Chai)

- tests: test/ceph\_test\_objectstore: make settings update and restore less error prone ([pr#21145](#), Igor Fedotov)
- tests: test: checking negative returns from creat() ([pr#18090](#), amitkuma)
- tests: test/CMakeLists: disable test\_pidfile.sh ([issue#20975](#), [pr#16977](#), Sage Weil)
- tests: test/CMakeLists: disable test-pidfile.sh ([pr#17401](#), Sage Weil)
- tests: test/coredumpctl: support freebsd ([pr#17447](#), Kefu Chai)
- tests: test/dashboard: hardcode .coverage path to workaround tox bugs ([pr#21485](#), Kefu Chai)
- tests: test/dashboard: specify workdir using tox.ini ([issue#23709](#), [pr#21416](#), Kefu Chai)
- tests: test: Don't dump core when using EXPECT\_DEATH ([pr#17390](#), Kefu Chai)
- tests: test/fio: extend fio objectstore plugin to better simulate OSD behavior ([pr#19101](#), Igor Fedotov)
- tests: test/fio: fix building of the fio\_ceph\_objectstore plugin ([pr#18332](#), Radoslaw Zarzynski)
- tests: test: Fix and enable test\_pidfile.sh ([issue#20770](#), [pr#16987](#), David Zafman)
- tests: test: Fix ceph-objectstore-tool usage check ([pr#17785](#), David Zafman)
- tests: test: fix misc fiemap testing ([issue#21716](#), [pr#18240](#), Kefu Chai, Ning Yao)
- tests: test: Initialization of \*comp\_racing\_read class CopyFromOp ([pr#17369](#), Amit Kumar)
- tests: test: Initializing ChunkReadOp members ([pr#19334](#), amitkuma)
- tests: test/journal: Initialize member variable m\_work\_queue ([pr#17089](#), amitkuma)
- tests: test/librados: be more tolerant with timed lock tests ([issue#20086](#), [pr#20161](#), Kefu Chai)
- tests: test/librados: increase pgp\_num along with pg\_num ([issue#23763](#), [pr#21555](#), Kefu Chai)
- tests: test/librados: s/invoke\_result\_t/result\_of\_t/ ([pr#20379](#), Kefu Chai)
- tests: test/librados\_test\_stub: pass snap context to zero op ([pr#17186](#), Mykola Golub)

- tests: test/log: fix for crash with libc++ ([pr#20233](#), Casey Bodley)
- tests: test: Make clearer by moving code out of loop ([pr#20759](#), David Zafman)
- tests: test/objectstore/test\_bluefs: cleanups ([pr#17909](#), Kefu Chai)
- tests: test: only test dashboard\_v2 when it is enabled ([pr#20777](#), Willem Jan Withagen)
- tests: test: only test enabled python bindings ([pr#21293](#), Kefu Chai)
- tests: test/osd: initialize Non-static class members in WeightedTestGenerator ([pr#15922](#), Jos Collin)
- tests: test/osd: Non-static class members not initialized in UnsetRedirectOp ([pr#15921](#), Jos Collin)
- tests: test/osd: silence warnings from -Wsign-compare ([pr#17027](#), Jos Collin)
- tests: test: put new BlueStore tests un ifdef WITH\_BLUESTORE ([pr#20576](#), Willem Jan Withagen)
- tests: test:qa:infra - Run update daily and use bash ([pr#21218](#), David Galloway)
- tests: test:qa:infra - teuthology crontab items as of 3/27/18 ([pr#21075](#), Yuri Weinstein)
- tests: test: reduce the chance to have degraded PGs ([issue#22711](#), [pr#20046](#), Kefu Chai)
- tests: test: remove distro\_version assert in distro detect test ([pr#21052](#), Shengjing Zhu)
- tests: test: Replace bc command with printf command ([pr#21013](#), David Zafman)
- tests: test: silence warning from -Wsign-compare ([pr#17790](#), Jos Collin)
- tests: test: silence warnings from -Wsign-compare ([pr#17962](#), Jos Collin)
- tests: tests - Replaced requests for "centos 7.3" to centos\_latest ([pr#19262](#), Yuri Weinstein)
- tests: test/store\_test: fix FTBFS as Sequencer is removed ([pr#20382](#), Kefu Chai)
- tests: test/store\_test: update Many4KWritesTest\* test cases to finalize with... ([pr#20230](#), Igor Fedotov)
- tests: test/throttle: kill tests exercising dtor of Throttle classes ([pr#17442](#), Kefu Chai)
- tests: test/unittest\_bufferlist: check retvals of syscalls ([pr#18238](#), Kefu Chai)

- tests: test/unittest\_pg\_log: silence gcc warning ([pr#17328](#), Kefu Chai)
- tests: test: Use jq in a compatible way and for easier diff analysis ([pr#21450](#), David Zafman)
- tests: test: Whitelist corrections ([pr#22167](#), David Zafman)
- tests,tools: crushtool: print error message to stderr not dout(1) ([issue#21758](#), [pr#18242](#), Kefu Chai)
- tests: unittest\_crypto: Don't exceed limit for getentropy ([pr#18505](#), Brad Hubbard)
- tests: vstart: fix initial start when there is no ceph.conf ([pr#21019](#), Jianpeng Ma)
- The Day Has Come! ([pr#19657](#), Adam C. Emerson)
- tools: Align use of uint64\_t in service\_daemon::AttributeType ([pr#16938](#), James Page)
- tools: ceph-disk: erase 110MB for nuking existing bluestore ([issue#22354](#), [pr#20400](#), Kefu Chai)
- tools: ceph-disk: fix '-runtime' omission for ceph-osd service ([issue#21498](#), [pr#17904](#), Carl Xiong)
- tools: ceph-disk: fix signed integer is greater than maximum when call major ([pr#19196](#), Song Shun)
- tools: ceph-disk: include output of failed command in exception ([pr#20497](#), Kefu Chai)
- tools: ceph-disk: more precise error message when a disk is specified ([pr#18018](#), Kefu Chai)
- tools: ceph-disk: reduce the scope of activate\_lock ([pr#20114](#), zhaokun)
- tools: ceph-disk: retry on OSError ([issue#21728](#), [pr#18162](#), Kefu Chai)
- tools: ceph-disk: unlock all partitions when activate ([pr#17363](#), Kefu Chai)
- tools: ceph-disk: write log to /var/log/ceph not to /var/run/ceph ([pr#18375](#), Kefu Chai)
- tools: ceph: fixes for "tell <service>.\*" command ([issue#21230](#), [pr#17463](#), Kefu Chai)
- tools: ceph-kvstore-tool: make it a bit more friendly ([pr#21477](#), Sage Weil)
- tools: ceph-kvstore-tool: use unique\_ptr<> to manage the lifecycle of bluestore ([pr#18221](#), Kefu Chai)

- tools: ceph-objectstore-tool: Add option “dump-import” to examine an export ([issue#22086](#), [pr#19368](#), David Zafman)
- tools: ceph-objectstore-tool: Fix set-size to clear data\_digest if changing ... ([pr#18885](#), David Zafman)
- tools: ceph-objectstore-tool: “\$OBJ get-omapdr” and “\$OBJ list-omap” scan all pgs instead of using specific pg ([issue#21327](#), [pr#17985](#), David Zafman)
- tools: ceph-objectstore-tool: skip object with generated version ([pr#18507](#), huangjun)
- tools: ceph-syn: silence clang analyzer warning ([pr#18577](#), Kefu Chai)
- tools: ceph-volume: Use a delimited CLI output parser instead of JSON ([pr#17097](#), Alfredo Deza)
- tools: cleanup: rip out ceph-rest-api ([issue#21264](#), [issue#22457](#), [pr#17530](#), Nathan Cutler)
- tools: correct total size formatting ([pr#21641](#), Zheng Yin)
- tools: crushtool: add -add-bucket and -move options ([pr#20183](#), Kefu Chai)
- tools: FreeBSD basic getopt does not use -options ([pr#21148](#), Willem Jan Withagen)
- tools: Initialization of \*server, command variables ([pr#17135](#), amitkuma)
- tools: make rados get/put/append command help txt clear ([issue#22958](#), [pr#20363](#), lvshuhua)
- tools: Modify “rados df” header’s alignment ([pr#17549](#), iliul)
- tools: rados add a cli option to clear omap keys ([issue#22255](#), [pr#19180](#), Abhishek Lekshmanan)
- tools: rados/tool: fixup rados stat command hint ([pr#16983](#), huanwen ren)
- tools: script: build-integration-branch: avoid Unicode error ([issue#24003](#), [pr#21918](#), Nathan Cutler)
- tools: script: ceph-release-notes: minor fixes for split\_component ([pr#16605](#), Abhishek Lekshmanan)
- tools: Special scrub handling of hinfo\_key errors ([issue#23428](#), [issue#23364](#), [pr#20947](#), David Zafman)
- tools: src/vstart.sh: default os to filestore for FreeBSD ([pr#17454](#), xie xingguo)
- tools: stop.sh: add ceph configure file location ([pr#20888](#), Jianpeng Ma)
- tools: tools/ceph-conf: dump parsed config in plain text or as json ([issue#21862](#),

[pr#18350](#), Piotr Dałek)

- tools: tools/ceph\_monstore\_tool: include mgrmap in initial paxos epoch ([issue#22266](#), [pr#19780](#), Kefu Chai)
- tools: tools/ceph\_monstore\_tool: rebuild initial mgrmap also ([issue#22266](#), [pr#19238](#), Kefu Chai)
- tools: tools/ceph-objectstore-tool: command to trim the pg log ([issue#23242](#), [pr#20786](#), Josh Durgin, David Zafman)
- tools: tools/ceph\_objectstore\_tool: fix 'dup' unable to duplicate meta PG ([pr#17572](#), xie xingguo)
- tools: tools/rados: improve the ls command usage ([pr#21553](#), Li Wang)
- tools: tools: rados: make -f be -format for consistency with ceph tool ([issue#15904](#), [pr#20147](#), Nathan Cutler)
- tools: tools/rados: use the monotonic clock in rados bench ([issue#21375](#), [pr#18588](#), Mohamad Gebai)
- tools: update monstore tool for fsmap, mgrmap ([issue#21577](#), [pr#18005](#), John Spray)
- tools: Use -no-mon-config so ceph\_objectstore\_tool.py test doesn't hang ([pr#21274](#), David Zafman)
- tools: vstart.sh: move rgw configuration to client.rgw section ([pr#18331](#), Yan Jun)
- tools: vstart.sh: use bluestore as default osd objectstore backend ([pr#17100](#), mychoxin)
- vstart: fix option (due to quotes) and allow disabling dashboard ([issue#23345](#), [pr#20986](#), Joao Eduardo Luis)
- vstart.sh: fix a typo ([pr#18729](#), iliul)
- vstart.sh: Fix help text in vstart.sh ([pr#21071](#), Marc Koderer)
- vstart.sh: quote cmd params when display executing cmd ([pr#17057](#), Jiaying Ren)
- vstart.sh: quote command only when necessary ([pr#18181](#), Kefu Chai)
- vstart.sh: should quote the parameters to get them quoted ([pr#18523](#), Kefu Chai)
- vstart.sh: simplify the objectstore related logic ([pr#17749](#), Kefu Chai)

## v12.2.13 Luminous

This is the 13th bug fix release of the Luminous v12.2.x long term stable release series. We recommend that all users upgrade to this release.

## Notable Changes

- Ceph now packages python bindings for python3.6 instead of python3.4, because EPEL7 recently switched from python3.4 to python3.6 as the native python3. see the announcement <<https://lists.fedoraproject.org/archives/list/epel-announce@lists.fedoraproject.org/message/EGUMKAIMPK2UD5VSHXM53BH2MBDGDWMO/>>\\_ for more details on the background of this change.
- We now have telemetry support via a ceph-mgr module. The telemetry module is absolutely on an opt-in basis, and is meant to collect generic cluster information and push it to a central endpoint. By default, we're pushing it to a project endpoint at <https://telemetry.ceph.com/report>, but this is customizable using by setting the 'url' config option with:

```
1. ceph telemetry config-set url '<your url>'
```

You will have to opt-in on sharing your information with:

```
1. ceph telemetry on
```

You can view exactly what information will be reported first with:

```
1. ceph telemetry show
```

Should you opt-in, your information will be licensed under the Community Data License Agreement - Sharing - Version 1.0, which you can read at <https://cdla.io/sharing-1-0/>

The telemetry module reports information about CephFS file systems, including:

- how many MDS daemons (in total and per file system)
- which features are (or have been) enabled
- how many data pools
- approximate file system age (year + month of creation)
- how much metadata is being cached per file system

As well as:

- whether IPv4 or IPv6 addresses are used for the monitors
- whether RADOS cache tiering is enabled (and which mode)
- whether pools are replicated or erasure coded, and which erasure code profile plugin and parameters are in use
- how many RGW daemons, zones, and zonegroups are present; which RGW frontends are in use
- aggregate stats about the CRUSH map, like which algorithms are used, how big buckets are, how many rules are defined, and what tunables are in use

- A health warning is now generated if the average osd heartbeat ping time exceeds a configurable threshold for any of the intervals computed. The OSD computes 1 minute, 5 minute and 15 minute intervals with average, minimum and maximum values. New configuration option `mon_warn_on_slow_ping_ratio` specifies a percentage of `osd_heartbeat_grace` to determine the threshold. A value of zero disables the warning. New configuration option `mon_warn_on_slow_ping_time` specified in milliseconds over-rides the computed value, causes a warning when OSD heartbeat pings take longer than the specified amount. New admin command `ceph daemon mgr.# dump_osd_network [threshold]` command will list all connections with a ping time longer than the specified threshold or value determined by the config options, for the average for any of the 3 intervals. New admin command `ceph daemon osd.# dump_osd_network [threshold]` will do the same but only including heartbeats initiated by the specified OSD.
- The configuration value `osd_calc_pg_upmaps_max_stddev` used for upmap balancing has been removed. Instead use the mgr balancer config `upmap_max_deviation` which now is an integer number of PGs of deviation from the target PGs per OSD. This can be set with a command like `ceph config set mgr mgr/balancer/upmap_max_deviation 2`. The default `upmap_max_deviation` is 1. There are situations where crush rules would not allow a pool to ever have completely balanced PGs. For example, if crush requires 1 replica on each of 3 racks, but there are fewer OSDs in 1 of the racks. In those cases, the configuration value can be increased.

## Changelog

- bluestore: >2GB bluefs writes ([pr#28965](#), kungf, Kefu Chai, Sage Weil)
- bluestore: Inspect allocations ([pr#29539](#), Neha Ojha, Adam Kupczyk)
- bluestore: [AFTER: #28644] luminous: os/bluestore: default to bitmap allocator for bluestore/bluefs ([pr#28972](#), Igor Fedotov)
- bluestore: add bluestore\_ignore\_data\_csum option ([pr#26247](#), Sage Weil)
- bluestore: apply shared\_alloc\_size to shared device with log level change ([pr#29910](#), Vikhyat Umrao, Josh Durgin, Igor Fedotov, Sage Weil)

- bluestore: avoid length overflow in extents returned by Stupid Alloc ([issue#40703](#), [pr#29025](#), Igor Fedotov)
- bluestore: call fault\_range properly prior to looking for blob to ... ([pr#27529](#), Igor Fedotov)
- bluestore: common/options: Set concurrent bluestore rocksdb compactions to 2 ([pr#30149](#), Mark Nelson)
- bluestore: dump before “no spanning blob id” abort ([pr#28030](#), Igor Fedotov)
- bluestore: fix assertion in StupidAllocator::get\_fragmentation ([pr#32523](#), Lei Liu, Igor Fedotov)
- bluestore: fix duplicate allocations in bmap allocator ([issue#40080](#), [pr#28644](#), Igor Fedotov)
- bluestore: fix improper setting of STATE\_KV\_SUBMITTED ([pr#31674](#), Igor Fedotov)
- bluestore: fix length overflow ([issue#39247](#), [pr#27365](#), Jianpeng Ma)
- bluestore: fix out-of-bound access in bmap allocator ([pr#27739](#), Igor Fedotov)
- bluestore: load OSD all compression settings unconditionally ([issue#40480](#), [pr#28895](#), Igor Fedotov)
- bluestore: os/bluestore/BitmapFreelistManager: disable bluestore\_debug\_freelist ([pr#27459](#), Sage Weil)
- bluestore: os/bluestore\_tool: bluefs-bdev-expand: indicate bypassed for main dev ([pr#27912](#), Igor Fedotov)
- bluestore: test/store\_test: fix/workaround for BlobReuseOnOverwriteUT and garbageCollection ([pr#27056](#), Igor Fedotov)
- build/ops: admin/build-doc: use python3 ([pr#30665](#), Kefu Chai, Jason Dillaman)
- build/ops: admin/build-doc: use python3 (follow-on fix) ([pr#30690](#), Nathan Cutler)
- build/ops: backport miscellaneous install-deps.sh and ceph.spec.in fixes from master ([issue#13997](#), [issue#37707](#), [issue#18163](#), [issue#22998](#), [pr#30722](#), Yao Guotao, Tomasz Setkowski, Andrey Parfenov, Alfredo Deza, Kefu Chai, Nathan Cutler, Yunchuan Wen, Zack Cerza, Brad Hubbard, Loic Dachary)
- build/ops: ceph-test RPM not built for SUSE ([pr#29736](#), Nathan Cutler)
- build/ops: cmake: pass -march to detect compiler support of arm64 crc/crypto ([issue#36080](#), [issue#17516](#), [pr#24169](#), Kefu Chai)
- build/ops: do\_cmake.sh: source not found ([issue#40004](#), [issue#39981](#), [pr#28216](#), Nathan Cutler)

- build/ops: install-deps.sh: Remove CR repo ([issue#13997](#), [pr#30129](#), Brad Hubbard, Alfredo Deza)
- build/ops: python-cephfs should depend on python-rados ([issue#37612](#), [issue#24918](#), [pr#27950](#), Kefu Chai)
- build/ops: python3-cephfs should provide python36-cephfs ([pr#30981](#), Kefu Chai)
- build/ops: rpm: Build with lttng on openSUSE ([issue#39332](#), [pr#27618](#), Nathan Cutler)
- build/ops: rpm: explicitly declare python-tox build dependency ([pr#31934](#), Nathan Cutler)
- ceph-volume: assume msgrV1 for all branches containing mimic ([pr#32796](#), Jan Fajerski)
- ceph-volume: batch functional idempotency test fails since message is now on stderr ([pr#29791](#), Jan Fajerski)
- ceph-volume: broken assertion errors after pytest changes ([pr#28929](#), Alfredo Deza)
- ceph-volume: do not fail when trying to remove crypt mapper ([pr#30556](#), Guillaume Abrioux)
- ceph-volume: does not recognize wal/db partitions created by ceph-disk ([pr#29462](#), Jan Fajerski)
- ceph-volume: fix stderr failure to decode/encode when redirected ([pr#30299](#), Alfredo Deza)
- ceph-volume: fix warnings raised by pytest ([pr#30677](#), Rishabh Dave)
- ceph-volume: lvm list is O(n^2) ([pr#30094](#), Rishabh Dave)
- ceph-volume: lvm.activate: Return an error if WAL/DB devices absent ([pr#29038](#), David Casier)
- ceph-volume: lvm.zap fix cleanup for db partitions ([issue#40664](#), [pr#30302](#), Dominik Csapak)
- ceph-volume: missing string substitution when reporting mounts ([issue#40978](#), [pr#29351](#), Shyukri Shyukriev)
- ceph-volume: pre-install python-apt and its variants before test runs ([pr#30296](#), Alfredo Deza)
- ceph-volume: prints errors to stdout with -format json ([issue#38548](#), [pr#29508](#), Jan Fajerski)
- ceph-volume: prints log messages to stdout ([pr#29603](#), Jan Fajerski, Kefu Chai,

Alfredo Deza)

- ceph-volume: set a lvm\_size property on the fakedevice fixture ([pr#30331](#), Andrew Schoen)
- ceph-volume: simple: when ‘type’ file is not present activate fails ([pr#29415](#), Alfredo Deza)
- ceph-volume: tests add a sleep in tox for slow OSDs after booting ([pr#28927](#), Alfredo Deza)
- ceph-volume: tests set the noninteractive flag for Debian ([pr#29901](#), Alfredo Deza)
- ceph-volume: update testing playbook ‘deploy.yml’ ([pr#29075](#), Andrew Schoen, Guillaume Abrioux)
- ceph-volume: use the Device.rotational property instead of sys\_api ([pr#28519](#), Andrew Schoen)
- ceph-volume: use the OSD identifier when reporting success ([pr#29771](#), Alfredo Deza)
- ceph-volume: zap always skips block.db, leaves them around ([issue#40664](#), [pr#30305](#), Alfredo Deza)
- cephfs: client: \_readdir\_cache\_cb() may use the readdir\_cache already clear ([issue#41148](#), [pr#30934](#), huanwen ren)
- cephfs: client: ceph.dir.rctime xattr value incorrectly prefixes 09 to the nanoseconds component ([issue#40166](#), [pr#28502](#), David Disseldorp)
- cephfs: client: clean up error checking and return of \_lookup\_parent ([issue#40085](#), [pr#28437](#), Jeff Layton)
- cephfs: client: return -EIO when sync file which unsafe reqs have been dropped ([issue#40877](#), [pr#30242](#), simon gao)
- cephfs: client: unlink dentry for inode with llref=0 ([issue#40960](#), [pr#29830](#), Xiaoxi CHEN)
- cephfs: kclient: nofail option not supported ([pr#28436](#), Kenneth Waegeman)
- cephfs: mds/server: check directory split after rename ([issue#39198](#), [issue#38994](#), [pr#27801](#), Shen Hang)
- cephfs: mds: add command that config individual client session ([issue#40811](#), [pr#31573](#), “Yan, Zheng”)
- cephfs: mds: add reference when setting Connection::priv to existing session ([pr#31049](#), “Yan, Zheng”)

- cephfs: mds: avoid trimming too many log segments after mds failover ([issue#40028](#), [pr#28543](#), simon gao)
- cephfs: mds: better output of 'ceph health detail' ([issue#39266](#), [pr#27848](#), Shen Hang)
- cephfs: mds: check dir fragment to split dir if mkdir makes it oversized ([pr#29829](#), Erqi Chen)
- cephfs: mds: cleanup truncating inodes when standby replay mds trim log segments ([pr#31286](#), "Yan, Zheng")
- cephfs: mds: dont print subtrees if they are too big or too many ([pr#27679](#), Rishabh Dave)
- cephfs: mds: drop reconnect message from non-existent session ([issue#39191](#), [issue#39026](#), [pr#27737](#), Shen Hang)
- cephfs: mds: fix corner case of replaying open sessions ([pr#28536](#), "Yan, Zheng")
- cephfs: mds: initialize cap\_revoke\_eviction\_timeout with conf ([issue#38844](#), [issue#39208](#), [pr#27840](#), simon gao)
- cephfs: mds: msg weren't destroyed before handle\_client\_reconnect returned, if the reconnect msg was from non-existent session ([issue#40588](#), [issue#40807](#), [pr#29097](#), Shen Hang)
- cephfs: mds: remove superfluous error in StrayManager::advance\_delayed() ([issue#38679](#), [pr#28432](#), "Yan, Zheng")
- cephfs: mds: reset heartbeat inside big loop ([pr#28544](#), "Yan, Zheng")
- cephfs: mds: there is an assertion when calling Beacon::shutdown() ([issue#38822](#), [pr#28438](#), huanwen ren)
- cephfs: mount: key parsing fail when doing a remount ([issue#40163](#), [pr#29226](#), Luis Henriques)
- cephfs: pybind/ceph\_volume\_client: remove ceph mds calls in favor of ceph fs calls ([issue#22038](#), [issue#22524](#), [pr#28445](#), Patrick Donnelly, Ramana Raja)
- cephfs: qa/cephfs: relax min\_caps\_per\_client check ([issue#38270](#), [issue#38686](#), [pr#27040](#), "Yan, Zheng")
- cephfs: qa: misc cache drop fixes ([issue#38340](#), [issue#38445](#), [pr#27342](#), Patrick Donnelly)
- common/config: hold lock while accessing mutable container ([pr#30345](#), Jason Dillaman)
- common: Keyrings created by ceph auth get are not suitable for ceph auth import

([issue#40548](#), [issue#22227](#), [pr#28742](#), Kefu Chai)

- common: common/ceph\_context: avoid unnecessary wait during service thread shutdown ([pr#31020](#), Jason Dillaman)
- common: common/options.cc: Lower the default value of osd\_deep\_scrub\_large omap\_object\_key\_threshold ([pr#29175](#), Neha Ojha)
- common: common/util: handle long lines in /proc/cpuinfo ([issue#38296](#), [pr#32349](#), Sage Weil)
- common: compressor/zstd: improvements ([pr#28647](#), Adam C. Emerson, Sage Weil)
- common: data race in OutputDataSocket ([issue#40188](#), [issue#40266](#), [pr#29202](#), Casey Bodley)
- core: ENOENT in collection\_move\_rename on EC backfill target ([issue#36739](#), [issue#38880](#), [pr#28110](#), Neha Ojha)
- core: Health warnings on long network ping times ([issue#40586](#), [issue#40640](#), [pr#30230](#), xie xingguo, David Zafman)
- core: Revert “crush: remove invalid upmap items” ([pr#32019](#), David Zafman)
- core: backport recent messenger fixes ([issue#39243](#), [issue#38242](#), [issue#39448](#), [pr#27583](#), xie xingguo, Jason Dillaman)
- core: ceph tell osd.xx bench help : gives wrong help ([issue#39006](#), [issue#39373](#), [pr#28112](#), Neha Ojha)
- core: ceph-objectstore-tool: rename dump-import to dump-export ([issue#39343](#), [issue#39284](#), [pr#27636](#), David Zafman)
- core: crc cache should be invalidated when posting preallocated rx buffers ([issue#38436](#), [pr#29248](#), Ilya Dryomov)
- core: crush/CrushWrapper: ensure crush\_choose\_arg\_map.size == max\_buckets ([issue#38664](#), [issue#38719](#), [pr#27085](#), Sage Weil)
- core: crush: remove invalid upmap items ([pr#31234](#), huangjun)
- core: lazy omap stat collection ([pr#29190](#), Brad Hubbard)
- core: mds,osd,mon,msg: use intrusive\_ptr for holding Connection::priv ([issue#20924](#), [pr#29859](#), Shinobu Kinjo, Kefu Chai, Jianpeng Ma, Samuel Just)
- core: mgr/localpool: pg\_num is an int arg to ‘osd pool create’ ([pr#30446](#), Sage Weil)
- core: mgr/prometheus: assign a value to osd\_dev\_node when obj\_store is not filestore or bluestore ([pr#31587](#), jiahuiweng)

- core: mon, osd: parallel clean\_pg\_upmaps ([issue#40229](#), [issue#40104](#), [pr#28594](#), xie xingguo)
- core: mon,osd: limit MOSDMap messages by size as well as map count ([issue#38276](#), [pr#28640](#), Sage Weil)
- core: mon/OSDMonitor: trim not-longer-exist failure reporters ([pr#30905](#), NancySu05)
- core: mon: Error message displayed when mon\_osd\_max\_split\_count would be exceeded is not as user-friendly as it could be ([issue#39353](#), [issue#39563](#), [pr#27908](#), Nathan Cutler, Brad Hubbard)
- core: mon: ensure prepare\_failure() marks no\_reply on op ([pr#30519](#), Joao Eduardo Luis)
- core: mon: mon/AuthMonitor: don't validate fs caps on authorize ([pr#28666](#), Joao Eduardo Luis)
- core: msg: output peer address when detecting bad CRCs ([issue#39367](#), [pr#27858](#), Greg Farnum)
- core: osd/OSDMap.cc: don't output over/underfull messages to lderr ([pr#31598](#), Neha Ojha)
- core: osd/OSDMap: Replace get\_out\_osds with get\_out\_existing\_osds ([issue#39154](#), [issue#39420](#), [pr#27728](#), Brad Hubbard)
- core: osd/OSDMap: do not trust partially simplified pg\_upmap\_item ([pr#30926](#), xie xingguo)
- core: osd/PG: Add PG to large omap log message ([pr#30922](#), Brad Hubbard)
- core: osd/PG: discover missing objects when an OSD peers and PG is degraded ([pr#27751](#), Jonas Jelten)
- core: osd/PGLog: preserve original\_crt to check rollbackability ([issue#38894](#), [issue#38905](#), [issue#36739](#), [issue#39042](#), [pr#27715](#), Neha Ojha)
- core: osd/PeeringState: recover\_got - add special handler for empty log ([pr#30896](#), xie xingguo)
- core: osd/PrimaryLogPG: skip obcs that don't exist during backfill scan\_range ([pr#31030](#), Sage Weil)
- core: osd/ReplicatedBackend.cc: 1321: FAILED assert(get\_parent()->get\_log().get\_log().objects.count(soid) && (get\_parent()->get\_log().get\_log().objects.find(soid)->second->op == pg\_log\_entry\_t::LOST\_REVERT) && (get\_parent()->get\_log().get\_log().object ([issue#39537](#), [issue#26958](#), [pr#28989](#), xie xingguo))

- core: osd/ReplicatedBackend.cc: 1349: FAILED ceph\_assert(peer\_missing.count(fromshard)) ([pr#31855](#), Neha Ojha, xie xingguo)
- core: osd/bluestore: Actually wait until completion in write\_sync ([pr#29564](#), Vitaliy Filippov)
- core: osd: Better error message when OSD count is less than osd\_pool\_default\_size ([issue#38617](#), [issue#38585](#), [pr#30298](#), Vikhyat Umrao, Kefu Chai, Sage Weil, zjh)
- core: osd: Diagnostic logging for upmap cleaning ([pr#32666](#), David Zafman)
- core: osd: FAILED ceph\_assert(attrs || !pg\_log.get\_missing().is\_missing(soid) || (it\_objects != pg\_log.get\_log().objects.end() && it\_objects->second->op == pg\_log\_entry\_t::LOST\_REVERT)) in PrimaryLogPG::get\_object\_context() ([issue#39218](#), [issue#38931](#), [issue#38784](#), [pr#27878](#), xie xingguo)
- core: osd: Fix for compatibility of encode/decode of osd\_stat\_t ([pr#31277](#), David Zafman)
- core: osd: Include dups in copy\_after() and copy\_up\_to() ([issue#39304](#), [pr#28185](#), David Zafman)
- core: osd: Remove unused osdmap flags full, nearfull from output ([issue#22350](#), [pr#30902](#), Gu Zhongyan, David Zafman)
- core: osd: add hdd, ssd and hybrid variants for osd\_snap\_trim\_sleep ([pr#31857](#), Neha Ojha)
- core: osd: clear PG\_STATE\_CLEAN when repair object ([pr#30271](#), Zengran Zhang)
- core: osd: fix out of order caused by letting old msg from down osd be processed ([pr#31293](#), Mingxin Liu)
- core: osd: merge replica log on primary need according to replica log's crt ([pr#30917](#), Zengran Zhang)
- core: osd: refuse to start if we're > N+2 from recorded require\_osd\_release ([issue#38076](#), [pr#31858](#), Sage Weil)
- core: osd: report omap/data/metadata usage ([issue#40638](#), [pr#28851](#), Sage Weil)
- core: osd: rollforward may need to mark pglog dirty ([issue#40403](#), [pr#31036](#), Zengran Zhang)
- core: osd: scrub error on big objects; make bluestore refuse to start on big objects ([pr#30785](#), Sage Weil, David Zafman)
- core: osd: shutdown recovery\_request\_timer earlier ([issue#39204](#), [pr#27810](#), Zengran Zhang)
- core: pybind: Rados.get\_fsid() returning bytes in python3 ([issue#38873](#),

- issue#38381, pr#27674, Jason Dillaman)
- core: should report EINVAL in ErasureCode::parse() if m<=0 (issue#38682, issue#38750, pr#28111, Sage Weil)
- doc: Minor rados related documentation fixes (issue#38896, issue#38902, pr#27185, David Zafman)
- doc: Missing Documentation for radosgw-admin reshard commands (man pages) (issue#40092, issue#21617, pr#28329, Orit Wasserman)
- doc: Update layout.rst (pr#26381, ypdai)
- doc: describe metadata\_heap cleanup (issue#18174, pr#30071, Dan van der Ster)
- doc: doc/rbd: s/guess/xml/ for codeblock lexer (pr#31091, Kefu Chai)
- doc: doc/rgw: document CreateBucketConfiguration for s3 PUT Bucket api (issue#39597, pr#31647, Casey Bodley)
- doc: doc/rgw: document use of 'realm pull' instead of 'period pull' (issue#39655, pr#30132, Casey Bodley)
- doc: fixed -read-only argument value in multisite doc (pr#31655, Chenjiong Deng)
- doc: osd\_recovery\_priority is not documented (but osd\_recovery\_op\_priority is) (pr#27471, David Zafman)
- doc: update bluestore cache settings and clarify data fraction (issue#39522, pr#31257, Jan Fajerski)
- doc: wrong datatype describing crush\_rule (pr#32267, Kefu Chai)
- doc: wrong value of usage log default in logging section (issue#37892, issue#37856, pr#29015, Abhishek Lekshmanan)
- mgr: Change default upmap\_max\_deviation to 5 (pr#32586, David Zafman)
- mgr: Release GIL and Balancer fixes (pr#31992, Kefu Chai, Noah Watkins, David Zafman)
- mgr: mgr/BaseMgrModule: drop GIL in set\_config (issue#39040, issue#36766, pr#27808, John Spray, xie xingguo, Sage Weil)
- mgr: mgr/balancer: blame if upmap won't actually work (issue#38781, pr#26498, xie xingguo)
- mgr: mgr/balancer: python3 compatibility issue (pr#31104, Mykola Golub)
- mgr: mgr/prometheus: Cast collect\_timeout (scrape\_interval) to float (pr#31107, Ben Meekhof)

- mgr: mgr/prometheus: replace whitespaces in metrics' names ([pr#31105](#), Alfonso Martínez)
- mgr: DaemonServer::handle\_conf\_change - broken locking ([issue#38899](#), [issue#38962](#), [pr#29213](#), xie xingguo)
- mgr: pybind/mgr: Cancel output color control ([pr#31696](#), Zheng Yin)
- mgr: restful: Query nodes\_by\_id for items ([pr#31272](#), Boris Ranto)
- mgr: telemetry module for mgr ([issue#37976](#), [pr#32135](#), Joao Eduardo Luis, Wido den Hollander, Kefu Chai, Sage Weil, Dan Mick)
- rbd: Reduce log level for cls/journal and cls/rbd expected errors ([issue#40865](#), [pr#30857](#), Jason Dillaman)
- rbd: journal: properly advance read offset after skipping invalid range ([pr#28811](#), Mykola Golub)
- rbd: krbd: avoid udev netlink socket overrun and retry on transient errors from udev\_enumerate\_scan\_devices() ([issue#39089](#), [pr#31360](#), Zhi Zhang, Ilya Dryomov)
- rbd: krbd: return -ETIMEDOUT in polling ([issue#38792](#), [issue#38975](#), [pr#27536](#), Dongsheng Yang)
- rbd: librbd: add missing shutdown states to managed lock helper ([issue#38387](#), [issue#38508](#), [pr#28158](#), Jason Dillaman)
- rbd: librbd: async open/close should free ImageCtx before issuing callback ([issue#39427](#), [issue#39031](#), [pr#28126](#), Jason Dillaman)
- rbd: librbd: disable image mirroring when moving to trash ([pr#28149](#), Mykola Golub)
- rbd: librbd: ensure compare-and-write doesn't skip compare after copyup ([issue#38440](#), [issue#38383](#), [pr#28134](#), Ilya Dryomov)
- rbd: librbd: improve object map performance under high IOPS workloads ([issue#38674](#), [issue#38538](#), [pr#28137](#), Jason Dillaman)
- rbd: librbd: properly track in-flight flush requests ([issue#40574](#), [pr#28773](#), Jason Dillaman)
- rbd: librbd: race condition possible when validating RBD pool ([issue#38500](#), [issue#38564](#), [pr#28140](#), Jason Dillaman)
- rbd: rbd-mirror: clear out bufferlist prior to listing mirror images ([issue#39460](#), [issue#39407](#), [pr#28124](#), Jason Dillaman)
- rbd: rbd-mirror: don't overwrite status error returned by replay ([pr#29874](#), Mykola Golub)

- rbd: rbd-mirror: handle duplicates in image sync throttler queue ([issue#40592](#), [issue#40519](#), [pr#28812](#), Mykola Golub)
- rbd: rbd-mirror: ignore errors relating to parsing the cluster config file ([pr#30118](#), Jason Dillaman)
- rbd: rbd-mirror: make logrotate work ([pr#32599](#), Mykola Golub)
- rbd: rbd/action: fix error getting positional argument ([issue#40095](#), [pr#29295](#), songweibin)
- rbd: tools/rbd-ggate: close log before running postfork ([pr#30858](#), Willem Jan Withagen)
- rbd: use the ordered throttle for the export action ([issue#40435](#), [pr#30856](#), Jason Dillaman)
- rgw: Adding tcp\_nodelay option to Beast ([issue#38925](#), [pr#27424](#), Or Friedmann)
- rgw: GetBucketCORS API returns Not Found error code when CORS configuration does not exist ([issue#38887](#), [issue#26964](#), [pr#27123](#), yuliyang, ashitakasam)
- rgw: LC: handle resharded buckets ([pr#29122](#), Abhishek Lekshmanan)
- rgw: RGWCoroutine::call(nullptr) sets retcode=0 ([pr#30329](#), Casey Bodley)
- rgw: TempURL should not allow PUTs with the X-Object-Manifest ([issue#20797](#), [pr#31652](#), Radoslaw Zarzynski)
- rgw: add list user admin OP API ([pr#30984](#), Oshyn Song)
- rgw: allow radosgw-admin to list bucket w -allow-unordered ([pr#31220](#), J. Eric Ivancich)
- rgw: civetweb frontend: response is buffered in memory if content length is not explicitly specified ([issue#39615](#), [issue#12713](#), [pr#28069](#), Robin H. Johnson)
- rgw: cls/rgw: raise debug level of bi\_log\_iterate\_entries output ([issue#40559](#), [pr#27974](#), Casey Bodley)
- rgw: cls/user: cls\_user\_set\_buckets\_info overwrites creation\_time ([issue#39635](#), [pr#31648](#), Casey Bodley)
- rgw: conditionally allow builtin users with non-unique email addresses ([issue#40089](#), [issue#40506](#), [pr#28717](#), Matt Benjamin)
- rgw: crypt: permit RGW-AUTO/default with SSE-S3 headers ([pr#31860](#), Matt Benjamin)
- rgw: datalog/mdlog trim commands loop until done ([pr#29713](#), Casey Bodley)
- rgw: delete\_obj\_index() takes mtime for bilog ([issue#24991](#), [pr#31649](#), Casey Bodley)

- rgw: don't crash on missing /etc/mime.types ([issue#38920](#), [issue#38328](#), [pr#27332](#), Casey Bodley)
- rgw: don't throw when accept errors are happening on frontend ([pr#30147](#), Yuval Lifshitz)
- rgw: failed to pass test\_bucket\_create\_naming\_bad\_punctuation in s3test ([issue#39360](#), [issue#39358](#), [issue#23587](#), [issue#26965](#), [pr#27668](#), yuliyang, Abhishek Lekshmanan)
- rgw: fix bucket may redundantly list keys after BI\_PREFIX\_CHAR ([issue#39984](#), [issue#40149](#), [pr#28408](#), Tianshan Qu, Casey Bodley)
- rgw: fix cls\_bucket\_list\_unordered() partial results ([pr#30254](#), Mark Kogan)
- rgw: fix drain handles error when deleting bucket with bypass-gc option ([pr#30198](#), dongdong tao)
- rgw: fix issue for CreateBucket with BucketLocation param ([pr#29826](#), Enming Zhang, Matt Benjamin)
- rgw: fix read not exists null version return wrong ([issue#38811](#), [issue#38908](#), [pr#27330](#), Tianshan Qu)
- rgw: fix refcount tags to match and update object's idtag ([pr#30323](#), J. Eric Ivancich)
- rgw: gc use aio ([issue#24592](#), [pr#28784](#), Yehuda Sadeh, Zhang Shaowen, Yao Zongyou, Jesse Williamson)
- rgw: get or set realm zonegroup zone need check user's caps ([issue#37497](#), [pr#28332](#), yuliang, Casey Bodley)
- rgw: housekeeping of reset stats operation in radosgw-admin and cls back-end ([pr#30674](#), J. Eric Ivancich)
- rgw: inefficient unordered bucket listing ([issue#39409](#), [issue#39393](#), [pr#28350](#), Casey Bodley)
- rgw: lc: continue past get\_obj\_state() failure ([pr#32194](#), Matt Benjamin)
- rgw: make dns hostnames matching case insensitive ([issue#40995](#), [pr#30375](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: mitigate bucket list with max-entries excessively high ([pr#30666](#), J. Eric Ivancich)
- rgw: multisite: 'radosgw-admin bucket sync status' should call syncs\_from(source.name) instead of id ([issue#40022](#), [issue#40143](#), [pr#29271](#), Casey Bodley)

- rgw: orphans find perf improvements ([issue#39180](#), [pr#28314](#), Abhishek Lekshmanan)
- rgw: parse\_copy\_location defers url-decode ([issue#27217](#), [pr#31651](#), Casey Bodley)
- rgw: policy fix for nonexistent objects ([issue#38638](#), [pr#29153](#), Pritha Srivastava)
- rgw: remove\_olh\_pending\_entries() does not limit the number of xattrs to remove ([issue#39118](#), [issue#39177](#), [pr#28349](#), Casey Bodley)
- rgw: resolve bugs and clean up garbage collection code ([issue#38454](#), [pr#31664](#), Dan Hill, J. Eric Ivancich)
- rgw: return ERR\_NO\_SUCH\_BUCKET early while evaluating bucket policy ([issue#38420](#), [pr#31218](#), Abhishek Lekshmanan)
- rgw: rgw-admin: fix data sync report for master zone ([issue#38958](#), [pr#27453](#), cfanz)
- rgw: rgw-admin: object stat command output's delete\_at not readable ([issue#39497](#), [pr#27991](#), Abhishek Lekshmanan)
- rgw: rgw/OutputDataSocket: actually discard data on full buffer ([issue#40178](#), [pr#31654](#), Matt Benjamin)
- rgw: rgw/multisite: Don't allow certain radosgw-admin commands to run on non-master zone ([issue#39548](#), [pr#30946](#), Danny Al-Gaaf, Shilpa Jagannath)
- rgw: rgw\_file: save etag and acl info in setattr ([issue#39227](#), [pr#27881](#), Tao Chen)
- rgw: rgw\_sync: drop ENOENT error logs from mdlog ([issue#40032](#), [issue#38748](#), [pr#27110](#), Abhishek Lekshmanan)
- rgw: set null version object acl issues ([issue#36763](#), [pr#31653](#), Tianshan Qu)
- rgw: the Multi-Object Delete operation of S3 API wrongly handles the Code response element ([issue#18241](#), [issue#40135](#), [pr#29269](#), Radoslaw Zarzynski)
- rgw: unable to cancel reshards operations for buckets with tenants ([issue#39016](#), [pr#27992](#), Abhishek Lekshmanan)
- rgw: update civetweb submodule to match version in mimic ([issue#24158](#), [pr#27982](#), Abhishek Lekshmanan)
- rgw: update s3-test download code for s3-test tasks ([pr#32227](#), Ali Maredia)
- rgw: when exclusive lock fails due existing lock, log add'l info ([issue#38397](#), [issue#38171](#), [pr#26554](#), J. Eric Ivancich)
- rgw: send x-amz-version-id header when upload files ([issue#39572](#), [pr#27935](#), Xinying Song)

- tools: Add clear-data-digest command to objectstore tool ([pr#29366](#), Li Yichao)
- tools: platform.linux\_distribution() is deprecated; stop using it ([issue#39277](#), [issue#18163](#), [pr#27557](#), Nathan Cutler)
- tools: rados tools list objects in a pg ([issue#36732](#), [pr#30608](#), Li Wang, Vikhyat Umrao)

## v12.2.12 Luminous

---

This is the twelfth bug fix release of the Luminous v12.2.x long term stable release series. We recommend that all users upgrade to this release.

## Notable Changes

---

- In 12.2.11 and earlier releases, keyring caps were not checked for validity, so the caps string could be anything. As of 12.2.12, caps strings are validated and providing a keyring with an invalid caps string to, e.g., ceph auth add will result in an error.

## Changelog

---

- auth: ceph auth add does not sanity-check caps ([issue#22525](#), [pr#24906](#), Jing Li, Nathan Cutler, Sage Weil)
- build/ops: Allow multi instances of “make tests” on the same machine ([issue#36737](#), [pr#26186](#), Kefu Chai)
- build/ops: rpm: require ceph-base instead of ceph-common ([issue#37620](#), [pr#25810](#), Sébastien Han)
- ceph-volume: add -all flag to simple activate ([pr#26656](#), Jan Fajerski)
- ceph-volume: look for rotational data in lsblk ([pr#26989](#), Andrew Schoen)
- ceph-volume: replace testinfra command with py.test ([pr#26824](#), Alfredo Deza)
- ceph-volume: revert partition as disk ([issue#37506](#), [pr#26295](#), Jan Fajerski)
- ceph-volume: simple scan will now scan all running ceph-disk OSDs ([pr#26857](#), Andrew Schoen)
- ceph-volume: use our own testinfra suite for functional testing ([pr#26703](#), Andrew Schoen)
- CLI: ability to change file ownership ([issue#38370](#), [pr#26758](#), Sébastien Han)
- client: session flush does not cause cap release message flush ([issue#38009](#),

[pr#26271](#), Patrick Donnelly)

- common: ceph\_timer: stop timer's thread when it is suspended ([issue#37766](#), [pr#26579](#), Peng Wang)
- common: fix for broken rbdmap parameter parsing ([issue#36327](#), [pr#26000](#), Marc Schoechlin)
- core: Objecter::calc\_op\_budget: Fix invalid access to extent union member ([issue#37932](#), [pr#26064](#), Simon Ruggier)
- core: os/filestore: ceph\_abort() on fsync(2) or fdatasync(2) failure ([issue#38258](#), [pr#26871](#), Sage Weil)
- crypto: don't use PK11\_ImportSymKey() in FIPS mode ([issue#38843](#), [pr#27104](#), Radoslaw Zarzynski)
- Fix recovery and backfill priority handling ([issue#27985](#), [issue#38041](#), [pr#26793](#), Sage Weil, xie xingguo, David Zafman)
- journal: max journal order is incorrectly set at 64 ([issue#37541](#), [pr#25955](#), Mykola Golub)
- librgw: export multitenancy support ([issue#37928](#), [pr#25986](#), Tao Chen)
- mds: broadcast quota message to client when disable quota ([issue#38054](#), [pr#26293](#), Junhui Tang)
- mds: fix potential re-evaluate stray dentry in \_unlink\_local\_finish ([issue#38263](#), [pr#26473](#), Zhi Zhang)
- mds: handle fragment notify race ([issue#36035](#), [pr#25990](#), "Yan, Zheng")
- mds: handle state change race ([issue#37594](#), [pr#26005](#), "Yan, Zheng")
- mds: log evicted clients to clog/dbg ([issue#37639](#), [pr#25858](#), Patrick Donnelly)
- mds: log new client sessions with various metadata ([issue#37678](#), [pr#26257](#), Patrick Donnelly)
- mds: message invalid access ([issue#38488](#), [pr#26661](#), Patrick Donnelly)
- MDSMonitor: do not assign standby-replay when degraded ([issue#36384](#), [pr#26642](#), Patrick Donnelly)
- MDSMonitor: missing osdmon writeable check ([issue#37929](#), [pr#26065](#), Patrick Donnelly)
- mds: optimize revoking stale caps ([issue#38043](#), [pr#26278](#), "Yan, Zheng")
- mds: stopping MDS with a large cache (40+GB) causes it to miss heartbeats ([issue#37723](#), [issue#38022](#), [pr#26232](#), Patrick Donnelly)

- mds: trim cache after journal flush ([issue#38010](#), [pr#26215](#), Patrick Donnelly)
- mds: wait for client to release shared cap when re-acquiring xlock ([issue#38491](#), [pr#27024](#), "Yan, Zheng")
- mds: wait shorter intervals if beacon not sent ([issue#36367](#), [pr#25979](#), Patrick Donnelly)
- mgr: "balancer execute" only requires read permissions ([issue#25345](#), [pr#25768](#), John Spray)
- mgr/balancer: restrict automatic balancing to specific weekdays ([pr#26501](#), xie xingguo)
- mgr/BaseMgrModule: drop GIL for ceph\_send\_command ([issue#38537](#), [pr#26830](#), Sage Weil)
- mgr/DaemonServer: log pgmap usage to cluster log ([issue#37886](#), [pr#26207](#), Neha Ojha)
- mgr/dashboard: fix for using '::' on hosts without ipv6 ([issue#38575](#), [pr#26751](#), Noah Watkins)
- mgr: deadlock: \_check\_auth\_rotating possible clock skew, rotating keys expired way too early ([issue#23460](#), [pr#26427](#), Yan Jun)
- mgr: drop GIL in StandbyPyModule::get\_config ([issue#35985](#), [pr#26613](#), John Spray; [pr#27639](#), wumingqiao)
- mgr/restful: fix py got exception when get osd info ([issue#38182](#), [pr#26199](#), Boris Ranto, zouaiguo)
- mon: A PG with PG\_STATE\_REPAIR doesn't mean damaged data, PG\_STATE\_IN... ([issue#38070](#), [pr#26305](#), David Zafman)
- mon/MgrStatMonitor: ensure only one copy of initial service map ([issue#38839](#), [pr#27207](#), Sage Weil)
- mon: monstore tool rebuild does not generate creating\_pgs ([issue#36306](#), [pr#25825](#), Sage Weil)
- mon: scrub warning check incorrectly uses mon scrub interval ([issue#37264](#), [pr#26557](#), Zhi Zhang, Sage Weil, David Zafman)
- msg/async: backport recent messenger fixes ([issue#36497](#), [issue#37778](#), [pr#25956](#), xie xingguo)
- msg/msg\_types: fix the dencoder of entity\_addr\_t ([issue#24676](#), [pr#26042](#), Kefu Chai)
- msgr: should set EPOLLET flag on del\_event() ([issue#38828](#), [pr#27226](#), Roman

Penyaev)

- msg: should set EPOLLET flag on del\_event() ([issue#38857](#), [pr#27226](#), Roman Penyaev)
- multisite: es sync null versioned object failed because of olh info ([issue#23842](#), [issue#23841](#), [pr#26358](#), Tianshan Qu, Shang Ding)
- Object can still be deleted even if s3:DeleteObject policy is set ([issue#37403](#), [pr#26310](#), Enming.Zhang)
- objecter: avoid race when reset down osd's session ([issue#24601](#), [pr#25853](#), Zengran Zhang)
- os/bluestore: backport new bitmap allocator ([issue#24598](#), [pr#26979](#), Radoslaw Zarzynski, Jianpeng Ma, Igor Fedotov, Sage Weil)
- os/bluestore: bitmap allocator might fail to return contiguous chunk despite having enough space ([issue#38761](#), [pr#27312](#), Igor Fedotov)
- os/bluestore: do not assert on non-zero err codes from compress() call ([issue#37839](#), [pr#26544](#), Igor Fedotov)
- os/bluestore: fix lack of onode ref during removal ([issue#38395](#), [pr#26540](#), Sage Weil)
- os/bluestore: Fix problem with bluefs's freespace not being balanced when kv\_sync\_thread is sleeping ([issue#38574](#), [pr#26866](#), Adam Kupczyk)
- os/bluestore: fixup access a destroy cond cause deadlock or undefined ([issue#37733](#), [pr#26261](#), linbing)
- os/bluestore: KernelDevice::read() does the EIO mapping now ([issue#36455](#), [pr#25855](#), Radoslaw Zarzynski)
- osd: backport recent upmap fixes ([issue#37968](#), [issue#37940](#), [issue#37881](#), [pr#26127](#), huangjun, xie xingguo)
- osd: backport recent upmap fixes ([issue#38826](#), [issue#38897](#), [pr#27224](#), huangjun, xie xingguo)
- osd/bluestore: deep fsck fails on inspecting very large onodes ([issue#38065](#), [pr#26387](#), Igor Fedotov)
- OSD crashes in get\_str\_map while creating with ceph-volume ([issue#38329](#), [pr#26900](#), Sage Weil)
- osd: keep using cache even if op will invalid cache ([issue#37593](#), [pr#26078](#), Zengran Zhang)
- osd/PG.cc: account for missing set irrespective of last\_complete ([issue#37919](#),

pr#26236, Neha Ojha)

- osd/PrimaryLogPG: fix the extent length error of the sync read ([issue#37680](#), pr#25711, Xiaofei Cui)
- osd/PrimaryLogPG: handle object !exists in handle\_watch\_timeout ([issue#38432](#), pr#26706, Sage Weil)
- rbd-mirror: update mirror status when stopping ([issue#36659](#), pr#25720, Jason Dillaman)
- rgw: bucket full sync handles delete markers ([issue#38007](#), pr#26192, Casey Bodley)
- rgw: bucket limit check misbehaves for > max-entries buckets (usually 1000) ([issue#35973](#), pr#26946, Matt Benjamin)
- rgw: bug in versioning concurrent, list and get have consistency issue ([issue#38060](#), pr#26548, Wang Hao)
- rgw: check for non-existent bucket in RGWGetACLs ([issue#38116](#), pr#26530, Matt Benjamin)
- rgw: data sync drains lease stack on lease failure ([issue#38479](#), pr#26761, Casey Bodley)
- rgw: fails to start on Fedora 28 from default configuration ([issue#24228](#), pr#26131, Matt Benjamin)
- rgw: feature - log successful bucket resharding events ([issue#37647](#), pr#25738, J. Eric Ivancich)
- rgw: fetch\_remote\_obj filters out olh attrs ([issue#37792](#), pr#26191, Casey Bodley)
- rgw: fix cls\_bucket\_head result order consistency ([issue#38410](#), pr#26546, Tianshan Qu)
- rgw: fix radosgw linkage with WITH\_RADOSGW\_BEAST\_FRONTEND=OFF ([issue#23680](#), pr#26332, Nathan Cutler)
- rgw: fix rgw\_data\_sync\_info::json\_decode() ([issue#38373](#), pr#26549, Casey Bodley)
- rgw: handle S3 version 2 pre-signed urls with meta-data ([issue#23470](#), pr#25901, Matt Benjamin)
- rgw: ldap: fix LDAPAuthEngine::init() when uri !empty() ([issue#38699](#), pr#27173, Matt Benjamin)
- rgw: multiple es related fixes and improvements ([issue#22877](#), [issue#23655](#), [issue#38030](#), [issue#38028](#), [issue#36092](#), pr#26516, Yehuda Sadeh, Abhishek Lekshmanan)

- rgw multisite: data sync checks empty next\_marker for datalog ([issue#39033](#), [pr#27299](#), Casey Bodley)
- rgw: nfs: skip empty (non-POSIX) path segments ([issue#38744](#), [pr#27180](#), Matt Benjamin)
- rgw: only update last\_trim marker on ENODATA ([issue#38075](#), [pr#26619](#), Casey Bodley)
- rgw: “radosgw-admin bucket rm ... -purge-objects” can hang ([issue#38007](#), [issue#38134](#), [pr#26263](#), J. Eric Ivancich)
- rgw: rgw\_file: only first subuser can be exported to nfs ([issue#37855](#), [pr#26677](#), MinSheng Lin)
- rgw: rgwgc: process coredump in some special case ([issue#23199](#), [pr#25611](#), zhaokun)
- rgw: sse-c-fixes ([issue#38700](#), [pr#27295](#), Adam Kupczyk, Casey Bodley, Abhishek Lekshmanan)
- rgw: sync module: avoid printing attrs of objects in log ([issue#37646](#), [pr#27030](#), Abhishek Lekshmanan)
- tools: ceph-objectstore-tool: Dump hashinfo ([issue#37597](#), [pr#25722](#), David Zafman)

## v12.2.11 Luminous

---

This is the eleventh bug fix release of the Luminous v12.2.x long term stable release series. We recommend that all users upgrade to this release. Please note the following precautions while upgrading.

## Notable Changes

---

- This release fixes the pg log hard limit bug that was introduced in 12.2.9, <https://tracker.ceph.com/issues/36686>. A flag called pglog\_hardlimit has been introduced, which is off by default. Enabling this flag will limit the length of the pg log. In order to enable that, the flag must be set by running ceph osd set pglog\_hardlimit after completely upgrading to 12.2.11. Once the cluster has this flag set, the length of the pg log will be capped by a hard limit. Once set, this flag *must not* be unset anymore.
- There have been fixes to RGW dynamic and manual resharding, which no longer leaves behind stale bucket instances to be removed manually. For finding and cleaning up older instances from a reshard a radosgw-admin command reshard stale-instances list and reshard stale-instances rm should do the necessary cleanup.
- cephfs-journal-tool makes rank argument (-rank) mandatory. Rank is of format

filesystem:rank, where filesystem is the CephFS filesystem and rank is the MDS rank on which the operation is to be executed. To operate on all ranks, use all or \* as the rank specifier. Note that, operations that dump journal information to file will now dump to per-rank suffixed dump files. Importing journal information from dump files is disallowed if operation is targeted for all ranks.

- CVE-2018-14662: mon: limit caps allowed to access the config store
- CVE-2018-16846: rgw: enforce bounds on max-keys/max-uploads/max-parts (issue#35994 <<http://tracker.ceph.com/issues/35994&gt;>>)
- CVE-2018-16889: rgw: sanitize customer encryption keys from log output in v4 auth (issue#37847 <<http://tracker.ceph.com/issues/37847&gt;>>)

## Changelog

---

- build/ops: cmake: link unittest\_compression against gtest ([pr#24921](#), Willem Jan Withagen)
- build/ops: run-make-check.sh ccache tweaks ([issue#24826](#), [issue#24817](#), [issue#24777](#), [pr#23902](#), Nathan Cutler, Erwan Velu)
- ceph-bluestore-tool: fix set label functionality for specific keys ([pr#25187](#), Igor Fedotov)
- ceph-create-keys: fix octal notation for Python 3 without losing compatibility with Python 2 ([issue#37643](#), [pr#25532](#), James Page)
- cephfs: ceph-volume-client: allow setting mode of CephFS volumes ([pr#25407](#), Tom Barron)
- cephfs-journal-tool: make -rank argument mandatory ([pr#24728](#), Venky Shankar)
- cephfs: mgr/status: fix fs status subcommand did not show standby-replay MDS' perf info ([issue#36575](#), [issue#36399](#), [pr#25032](#), Zhi Zhang)
- cephfs: race of updating wanted caps ([issue#37635](#), [issue#37464](#), [pr#25762](#), "Yan, Zheng")
- ceph-volume: Adapt code to support Python3 ([pr#26030](#), Volker Theile)
- ceph-volume add device\_id to inventory listing ([pr#25350](#), Jan Fajerski)
- ceph-volume: enable device discards ([issue#36532](#), [pr#25748](#), Jonas Jelten)
- ceph-volume: fix Batch object in py3 environments ([pr#25552](#), Jan Fajerski)
- ceph-volume: fix JSON output in inventory ([issue#37390](#), [pr#25922](#), Sebastian Wagner)

- ceph-volume: Fix TypeError: join() takes exactly one argument (2 given) ([issue#37595](#), [pr#25772](#), Sebastian Wagner)
- ceph-volume fix TypeError on dmcrypt when using Python3 ([pr#26114](#), Alfredo Deza)
- ceph-volume: introduce class hierarchy for strategies ([pr#25553](#), Jan Fajerski, Alfredo Deza)
- ceph-volume: mark a device not available if it belongs to ceph-disk ([pr#26117](#), Andrew Schoen)
- ceph-volume normalize comma to dot for string to int conversions ([issue#37442](#), [pr#25776](#), Alfredo Deza)
- ceph-volume: set permissions right before prime-osd-dir ([issue#37486](#), [pr#25778](#), Andrew Schoen, Alfredo Deza)
- ceph-volume tests/functional declare ceph-ansible roles instead of importing them ([issue#37805](#), [pr#25838](#), Alfredo Deza)
- ceph-volume zap devices associated with an OSD ID and/or OSD FSID ([pr#26014](#), Alfredo Deza)
- ceph-volume: zap: improve zapping to remove all partitions and all LVs, encrypted or not ([issue#37449](#), [pr#25352](#), Alfredo Deza)
- cli: dump osd-fsid as part of osd find <id> ([issue#37966](#), [pr#26036](#), Noah Watkins)
- client: do not move f->pos until success write ([issue#37631](#), [pr#25684](#), Junhui Tang)
- client: explicitly show blacklisted state via asok status command ([issue#36456](#), [issue#36352](#), [pr#24994](#), Jonathan Brielmaier, Zhi Zhang)
- client: fix fuse client hang because its pipe to mds is not ok4 ([issue#37829](#), [issue#36079](#), [pr#25904](#), Guan yunfei)
- client: request next osdmap for blacklisted client ([issue#36668](#), [issue#36691](#), [pr#24986](#), Zhi Zhang)
- common: auth/AuthSessionHandler: no handler if no session key ([issue#37427](#), [issue#36443](#), [pr#25297](#), Sage Weil)
- common/blkdev, ceph-volume: improve get\_device\_id ([pr#25752](#), Sage Weil)
- common: fix memory leaks in WeightedPriorityQueue ([issue#37429](#), [issue#36248](#), [pr#25296](#), Radoslaw Zarzynski)
- common: (mon) command sanitization accepts floats when Int type is defined resulting in exception fault in ceph-mon ([issue#26919](#), [pr#24374](#), Sage Weil)
- common: shut up some warnings ([pr#24648](#), Kefu Chai)

- config: drop config::lock when invoking config observer ([issue#37762](#), [pr#25833](#), Kefu Chai, Venky Shankar)
- core: bluestore: rename does not old ref to replacement onode at old name ([issue#36541](#), [issue#36638](#), [pr#24989](#), Jonathan Brielmair, Sage Weil)
- core: enable the pg deletion process to be throttled ([issue#36321](#), [pr#24501](#), David Zafman)
- core: mgr crash on scrub of unconnected osd ([issue#36110](#), [issue#36464](#), [pr#25030](#), Sage Weil)
- core: mon osdmap cash too small during upgrade to mimic ([issue#36506](#), [pr#25021](#), Sage Weil)
- core: Objecter: add ignore cache flag if got redirect reply ([issue#36657](#), [pr#25074](#), Iain Buclaw, Jonathan Brielmair)
- core: os/bluestore\_tool: fix bluefs expand ([pr#25384](#), Igor Fedotov)
- core: rados rm -force-full is blocked when cluster is in full status ([issue#36436](#), [pr#25018](#), Yang Honggang)
- crushtool: add -reclassify operation to convert legacy crush maps to use device classes ([pr#25307](#), Sage Weil)
- debian: correct ceph-common relationship with older radosgw package ([pr#24997](#), Matthew Vernon)
- doc: broken link on troubleshooting-mon page ([pr#25500](#), James McClune)
- doc: fix broken fstab url in cephfs/fuse ([issue#36286](#), [pr#24434](#), Jos Collin)
- doc: Fix typo error on cephfs/fuse/ ([issue#36180](#), [issue#36309](#), [pr#24752](#), Karun Josy)
- doc: Put command template into literal block ([pr#25001](#), Alexey Stupnikov)
- doc/rados: update bluestore provisioning and autotuning docs ([issue#37341](#), [pr#25284](#), Mark Nelson)
- doc: show edit on github links and version warnings ([pr#25267](#), Neha Ojha, Noah Watkins)
- doc/user-management: Remove obsolete reset caps command ([issue#37663](#), [pr#25609](#), Brad Hubbard)
- examples: fix link order in librados example Makefile ([issue#37795](#), [pr#25829](#), Mahati Chamarthi)
- extend reconnect period when mds is busy ([issue#37739](#), [pr#25784](#), "Yan, Zheng")

- fsck: cid is improperly matched to oid ([issue#36145](#), [issue#32731](#), [pr#24705](#), Sage Weil)
- libcephfs: expose CEPH\_SETATTR\_MTIME\_NOW and CEPH\_SETATTR\_ATIME\_NOW ([issue#36206](#), [issue#35961](#), [pr#24465](#), Zhu Shangzhong)
- librbd: fix missing unblock\_writes if shrink is not allowed ([issue#37363](#), [issue#36778](#), [pr#25253](#), runsisi)
- librbd: reset snaps in rbd\_snap\_list() ([issue#37535](#), [issue#37508](#), [pr#25458](#), Kefu Chai)
- mds: add “drop cache” command ([issue#36695](#), [issue#36281](#), [pr#24468](#), Rishabh Dave, Patrick Donnelly, Venky Shankar)
- mds: clean up log messages for standby-replay ([pr#25804](#), Patrick Donnelly)
- mds: create heartbeat grace config option ([issue#37674](#), [issue#37820](#), [pr#25889](#), Patrick Donnelly)
- mds: directories pinned keep being replicated back and forth between exporting mds and importing mds ([issue#37368](#), [issue#37606](#), [pr#25522](#), Xuehan Xu)
- mds: disallow dumping huge caches to formatter ([issue#37608](#), [pr#25567](#), Venky Shankar)
- mds: do not call Journaler::\_trim twice ([issue#37566](#), [issue#37629](#), [pr#25562](#), Tang Junhui)
- mds: fix bug filelock stuck at LOCK\_XSYN leading client can't read data ([issue#37700](#), [issue#37333](#), [pr#25677](#), Guan yunfei)
- mds: fix incorrect l\_pq\_executing\_ops statistics when meet an invalid item in purge queue ([issue#37627](#), [issue#37567](#), [pr#25560](#), Junhui Tang)
- mds: fix infinite loop in OpTracker::check\_ops\_in\_flight ([issue#37977](#), [pr#26048](#), “Yan, Zheng”)
- mds: fix infinite loop in OpTracker::check\_ops\_in\_flight ([issue#37977](#), [pr#26088](#), “Yan, Zheng”)
- mds: fix mds damaged due to unexpected journal length ([issue#36200](#), [pr#24440](#), Zhi Zhang)
- mds: migrate strays part by part when shutdown mds ([issue#26926](#), [issue#32091](#), [pr#24324](#), “Yan, Zheng”)
- MDSMonitor: allow beacons from stopping MDS that was laggy ([issue#37737](#), [pr#25686](#), Patrick Donnelly)
- mds: obsolete MDSMap option configs ([issue#37540](#), [pr#25431](#), Patrick Donnelly)

- mds: purge queue recovery hangs during boot if PQ journal is damaged ([issue#37899](#), [issue#37543](#), [pr#25968](#), Patrick Donnelly)
- mds: PurgeQueue write error handler does not handle EBLACKLISTED ([issue#37604](#), [pr#25524](#), Patrick Donnelly)
- mds: rctime not set on system inode (root) at startup ([issue#36221](#), [issue#36460](#), [pr#25043](#), Patrick Donnelly)
- mds: remove duplicated l\_mdc\_num\_strays perfcounter set ([issue#37633](#), [issue#37516](#), [pr#25682](#), Zhi Zhang)
- mds: severe internal fragment when decoding xattr\_map from log event ([issue#37399](#), [issue#37602](#), [pr#25520](#), "Yan, Zheng")
- mds: "src/mds/MDLog.cc: 281: FAILED ceph\_assert(!capped)" during max\_mds thrashing ([issue#36350](#), [issue#37092](#), [pr#25826](#), "Yan, Zheng")
- mgr/balancer: add cmd to list all plans ([issue#37420](#), [pr#25259](#), Yang Honggang)
- mgr/balancer: add crush\_compat\_metrics param ([issue#37413](#), [pr#25257](#), Dan van der Ster)
- mgr: balancer: python 3 compat fixes ([issue#37416](#), [pr#25258](#), Noah Watkins)
- mgr: fix crash due to multiple sessions from daemons with same name ([pr#25867](#), Mykola Golub)
- mgr: hold lock while accessing the request list and submitting request ([pr#25047](#), Jerry Lee)
- mgr: Module 'influx' has failed ([issue#25201](#), [pr#25184](#), Nathan Cutler, Wido den Hollander)
- mgr: prometheus: added bluestore db and wal devices to ceph\_disk\_occupation metric.// ([issue#37362](#), [pr#25216](#), Konstantin Shalygin)
- mgr: race between daemon state and service map in 'service status' ([issue#37478](#), [issue#36656](#), [pr#25369](#), Mykola Golub)
- mgr: [restful] deep\_scrub is not a valid OSD command ([issue#36720](#), [issue#36750](#), [pr#25041](#), Boris Ranto)
- mon: mark REMOVE\_SNAPS messages as no\_reply ([issue#37568](#), [issue#37694](#), [pr#25779](#), "Yan, Zheng")
- mon/OSDMonitor: do not populate void pg\_temp into nextmap ([issue#37811](#), [pr#25845](#), Aleksei Zakharov)
- mon: shutdown messenger early to avoid accessing deleted logger ([issue#37780](#), [issue#37813](#), [pr#25847](#), ningtao)

- os/bluestore: avoid frequent allocator dump on bluefs rebalance failure ([pr#24543](#), Igor Fedotov)
- os/bluestore/BlueStore.cc: 1025: FAILED assert(buffer\_bytes >= b->length) from ObjectStore/StoreTest.ColSplitTest2/2 ([issue#26943](#), [issue#24439](#), [pr#24992](#), Jonathan Brielmaier, Sage Weil)
- os/bluestore: handle spurious read errors ([issue#22464](#), [pr#24649](#), Paul Emmerich)
- osd: backport recent upmap fixes ([pr#25418](#), ningtao, xie xingguo)
- osdc/Objecter: update op\_target\_t::paused in \_calc\_target ([issue#37398](#), [issue#37553](#), [pr#25719](#), Song Shun, runsisi)
- osdc: reduce ObjectCacher's memory fragments ([issue#36642](#), [issue#36192](#), [pr#24872](#), "Yan, Zheng")
- osd: failed assert when osd\_memory\_target options mismatch ([issue#37697](#), [issue#37507](#), [pr#25604](#), xie xingguo)
- osd/mon: pg log hard limit with upgrades fixed ([issue#37903](#), [issue#21416](#), [pr#25949](#), Neha Ojha, xie xingguo)
- osd/OSD.cc: log slow requests in OSD logs ([pr#25824](#), Neha Ojha)
- osd/OSDMap: cancel mapping if target osd is out ([issue#37501](#), [pr#25698](#), ningtao, xie xingguo)
- osd: potential deadlock in PG::\_scan\_snaps when repairing snap mapper ([issue#36630](#), [pr#24833](#), Mykola Golub)
- osd: Prioritize user specified scrubs ([issue#37343](#), [issue#37269](#), [pr#25514](#), kungf, David Zafman)
- osd: race condition opening heartbeat connection ([issue#36602](#), [issue#36636](#), [pr#25035](#), Sage Weil)
- osd: RBD client IOPS pool stats are incorrect (2x higher; includes IO hints as an op) ([issue#24909](#), [issue#36556](#), [pr#25025](#), Jason Dillaman)
- pybind/mgr/status: fix ceph fs status in py3 environments ([issue#37573](#), [issue#37625](#), [pr#25695](#), Jan Fajerski)
- rbd: pybind: added missing RBD\_FLAG\_FAST\_DIFF\_INVALID constant ([issue#36407](#), [pr#25006](#), Jason Dillaman)
- rbd: [rbd-mirror] periodic mirror status timer might fail to be scheduled ([issue#36500](#), [issue#36554](#), [pr#24917](#), Nathan Cutler, Jason Dillaman)
- rgw: add ssl support to beast frontend ([issue#22832](#), [issue#24358](#), [pr#24621](#), Casey Bodley)

- rgw: apply quota config to users created via external auth ([issue#24595](#), [issue#36222](#), [pr#24547](#), Casey Bodley, Matt Benjamin)
- rgw: beast frontend fails to parse ipv6 endpoints ([issue#36733](#), [issue#36662](#), [pr#25512](#), Casey Bodley)
- rgw: bucket resharding fixes ([issue#37446](#), [issue#36688](#), [pr#25326](#), Orit Wasserman, Abhishek Lekshmanan, J. Eric Ivancich)
- rgw: catch exceptions from librados::NObjectIterator ([issue#37091](#), [issue#37475](#), [pr#25289](#), Casey Bodley)
- rgw: Don't treat colons specially in resource part of ARN ([issue#37482](#), [issue#23817](#), [pr#25387](#), Adam C. Emerson)
- rgw: es fixes for working with nfs ganesha ([issue#37349](#), [issue#36233](#), [issue#22758](#), [pr#25444](#), Abhishek Lekshmanan)
- rgw\_file: user info never synced since librgw init ([issue#37549](#), [pr#25484](#), Tao Chen)
- rgw: fixes for zone deletion ([issue#37328](#), [issue#37466](#), [pr#25320](#), Abhishek Lekshmanan)
- rgw: fix max-size in radosgw-admin and REST Admin API ([issue#37519](#), [pr#25448](#), Nick Erdmann)
- rgw: fix version bucket stats ([issue#37563](#), [issue#21429](#), [pr#25644](#), Shasha Lu)
- rgw: librgw: crashes in multisite configuration ([issue#36302](#), [issue#36414](#), [pr#24909](#), Casey Bodley)
- rgw: multisite: sync gets stuck retrying deletes that fail with ERR\_PRECONDITION\_FAILED ([issue#37551](#), [issue#37448](#), [pr#25506](#), Casey Bodley)
- rgw: radosgw-admin: translate reshard status codes (trivial) ([issue#37284](#), [issue#36486](#), [pr#25195](#), Matt Benjamin)
- rgw: rgw-admin: reshard add can add a non-existent bucket ([issue#36449](#), [issue#36757](#), [pr#25088](#), Jonathan Brielmair, Abhishek Lekshmanan)
- rgw: SSE encryption does not detect ssl termination in proxy ([issue#36644](#), [issue#27221](#), [pr#24944](#), Jonathan Brielmair, Casey Bodley)
- rpm: Use hardened LDFLAGS ([issue#36316](#), [issue#36391](#), [pr#25173](#), Boris Ranto)

## v12.2.10 Luminous

---

This is the tenth bug fix release of the Luminous v12.2.x long term stable release series. The previous release, v12.2.9, introduced the PG hard-limit patches which were

found to cause an issue in certain upgrade scenarios, and this release was expedited to revert those patches. If you already successfully upgraded to v12.2.9, you should **not** upgrade to v12.2.10, but rather **wait** for a release in which <http://tracker.ceph.com/issues/36686> is addressed. All other users are encouraged to upgrade to this release.

## Notable Changes

---

### OSD

- This release reverts the PG hard-limit patches added in v12.2.9.

## Changelog

---

- ceph-volume: add some choose\_disk capabilities ([issue#36446](#), [pr#24783](#), Erwan Velu)
- ceph-volume: remove version reporting from help menu ([issue#36386](#), [pr#24754](#), Alfredo Deza)
- ceph-volume: systemd import main so console\_scripts work for executable ([issue#36648](#), [pr#24853](#), Alfredo Deza)
- ceph-volume: tests install ceph-ansible's requirements.txt dependencies ([issue#36672](#), [pr#24960](#), Alfredo Deza)
- ceph-volume: util.encryption don't push stderr to terminal ([issue#36246](#), [pr#24827](#), Alfredo Deza)
- ceph-volume: util.encryption robust blkid+lsblk detection of lockbox ([pr#24981](#), Alfredo Deza)
- ceph-volume: use console\_scripts ([issue#36601](#), [pr#24837](#), Mehdi Abaakouk)
- OSDMapMapping does not handle active.size() > pool size ([issue#26866](#), [issue#35935](#), [pr#24432](#), Sage Weil)
- PG: add custom\_reaction Backfilled and release reservations ([issue#24333](#), [pr#23493](#), Neha Ojha)
- Revert "PG: add custom\_reaction Backfilled and release reservations after backfill" ([pr#24902](#), Neha Ojha)
- Revert pg log limit changes ([issue#36686](#), [pr#24903](#), Neha Ojha)
- backport and other test fixes for osd-scrub-repair.sh ([issue#35845](#), [issue#36393](#), [pr#24532](#), Xinying Song, David Zafman)
- ceph-volume tests.systemd update imports for systemd module ([issue#36704](#),

pr#24958, Alfredo Deza)

- ceph-volume: adds a --prepare flag to lvm batch ([issue#36363](#), [pr#24759](#), Andrew Schoen)
- cls/user: cls\_user\_remove\_bucket writes modified header ([issue#36534](#), [issue#36496](#), [pr#24855](#), Casey Bodley)
- core: bypass cache if performing deep scrub ([issue#35067](#), [pr#24802](#), Xiaoguang Wang)
- crush/CrushWrapper: fix crush tree json dumper ([issue#36149](#), [pr#24482](#), Oshyn Song)
- ec: src/common/interval\_map.h: 161: FAILED assert(len > 0) ([issue#21931](#), [issue#22330](#), [pr#24582](#), Neha Ojha)
- gperftools-libs-2.6.1-1 or newer required for binaries linked against corresponding version at build time ([issue#36552](#), [issue#23657](#), [issue#36558](#), [issue#36508](#), [pr#24706](#), Brad Hubbard)
- osd: add creating to pg\_string\_state ([issue#36174](#), [issue#36297](#), [pr#24602](#), Dan van der Ster)
- osd: cast 'whoami' to unsigned so it can be used as the seed for RNG ([issue#26890](#), [pr#24659](#), Kefu Chai)
- osdc/Objecter: possible race condition with connection reset ([issue#36183](#), [issue#36295](#), [pr#24574](#), Jason Dillaman)
- qa: add test that builds example librados programs ([issue#36229](#), [issue#15100](#), [pr#24538](#), Nathan Cutler)
- qa/ceph-ansible: Specify stable-3.2 branch ([issue#37331](#), [pr#25170](#), Brad Hubbard)
- rgw/beast: drop privileges after binding ports ([issue#36041](#), [pr#24454](#), Paul Emmerich)
- rgw: RGWAsyncGetBucketInstanceInfo does not access coroutine memory ([issue#35812](#), [issue#36212](#), [pr#24507](#), Casey Bodley)
- rgw: fix leak of curl handle on shutdown ([issue#35715](#), [issue#36214](#), [pr#24519](#), Casey Bodley)
- rgw: list bucket can not show the object uploaded by RGWPostObj when enable bucket versioning ([pr#24570](#), yuliyang)
- rgw: multisite: enforce spawn\_window for data full sync ([issue#26897](#), [pr#24857](#), Casey Bodley)

- rgw: set default objecter\_inflight\_ops = 24576 ([issue#36570](#), [issue#25109](#), [pr#24862](#), Matt Benjamin)
- rgw: user stats account for resharded buckets ([pr#24854](#), Casey Bodley)
- segv in BlueStore::OldExtent::create ([issue#36526](#), [issue#36591](#), [pr#24746](#), Sage Weil)
- test/common: unittest\_mclock\_priority\_queue builds with “make” command ([pr#24808](#), J. Eric Ivancich)

## v12.2.9 Luminous

---

This is the ninth bug fix release of the Luminous v12.2.x long term stable release series. Although this release contains several bugfixes across all the components, it also introduced the PG hard-limit patches which could cause problems during upgrade when not all PGs were active+clean. Therefore, users should not install this release. Instead, they should skip it and upgrade to 12.2.10 directly.

## Notable Changes

---

### OSD

- 12.2.9 contains the pg hard limit patches (<https://tracker.ceph.com/issues/23979>). A partial upgrade during recovery/backfill, can cause the osds on the previous version, to fail with assert(trim\_to <= info.last\_complete). The workaround for users is to upgrade and restart all OSDs to a version with the pg hard limit, or only upgrade when all PGs are active+clean. This patch will be reverted in 12.2.10, until a clean upgrade path is added to the pg log hard limit patches.

See also: <http://tracker.ceph.com/issues/36686>

- The bluestore\_cache\_\* options are no longer needed. They are replaced by osd\_memory\_target, defaulting to 4GB. BlueStore will expand and contract its cache to attempt to stay within this limit. Users upgrading should note this is a higher default than the previous bluestore\_cache\_size of 1GB, so OSDs using BlueStore will use more memory by default.

For more details, see BlueStore docs

[\\\_](http://docs.ceph.com/docs/master/rados/configuration/bluestore-config-ref/#cache-size)

## Changelog

---

- build/ops: add e2fsprogs runtime dependency ([pr#24663](#), Guillaume Abrioux, Alfredo

Deza)

- build/ops: deb: fix ceph-mgr .pyc files left behind ([issue#26883](#), [pr#23832](#), Dan Mick)
- build/ops: deb: require fuse for ceph-fuse ([issue#21057](#), [pr#23693](#), Thomas Serlin)
- build/ops: rpm: selinux-policy fixes ([pr#24136](#), Brad Hubbard)
- build/ops: rpm: use updated gperftools ([issue#35969](#), [pr#24259](#), Kefu Chai)
- ceph-volume: activate option -auto-detect-objectstore respects -no-systemd ([issue#36249](#), [pr#24358](#), Alfredo Deza)
- ceph-volume: lsblk can fail to find PARTLABEL, must fallback to blkid ([issue#36098](#), [pr#24335](#), Alfredo Deza)
- ceph-volume: add new ceph-handlers role from ceph-ansible ([issue#36251](#), [pr#24338](#), Alfredo Deza)
- ceph-volume: batch carve out lvs for bluestore ([issue#34535](#), [pr#24075](#), Alfredo Deza)
- ceph-volume: batch tests for mixed-type of devices ([issue#35535](#), [issue#27210](#), [pr#23967](#), Alfredo Deza)
- ceph-volume: batch: allow -osds-per-device, default it to 1 ([issue#35913](#), [pr#24080](#), Alfredo Deza)
- ceph-volume: batch: allow journal+block.db sizing on the CLI ([issue#36088](#), [pr#24209](#), Alfredo Deza)
- ceph-volume: custom cluster names fail on filestore trigger ([issue#27210](#), [pr#24280](#), Alfredo Deza)
- ceph-volume: do not send (lvm) stderr/stdout to the terminal, use the logfile ([issue#36492](#), [pr#24741](#), Alfredo Deza)
- ceph-volume: earlier detection for -journal and -filestore flag requirements ([issue#24794](#), [pr#24206](#), Alfredo Deza)
- ceph-volume: fix journal and filestore data size in lvm batch -report ([issue#36242](#), [pr#24307](#), Andrew Schoen)
- ceph-volume: fix zap not working with LVs ([issue#35970](#), [pr#24082](#), Alfredo Deza)
- ceph-volume: lvm.prepare update help to indicate partitions are needed, not devices ([issue#24795](#), [pr#24451](#), Jeffrey Zhang, Alfredo Deza)
- ceph-volume: make lvm batch idempotent ([pr#24589](#), Andrew Schoen)
- ceph-volume: remove version reporting from help menu ([issue#36386](#), [pr#24754](#),

Alfredo Deza)

- ceph-volume: skip processing devices that don't exist when scanning system disks ([issue#36247](#), [pr#24382](#), Alfredo Deza)
- cephfs: MDSMonitor: consider raising priority of MMDSBeacons from MDS so they are processed before other client messages ([issue#26899](#), [pr#23554](#), Patrick Donnelly)
- cephfs: MDSMonitor: lookup of gid in prepare\_beacon that has been removed will cause exception ([issue#35848](#), [pr#23990](#), Patrick Donnelly)
- cephfs: ceph-fuse: add SELinux policy ([issue#36103](#), [pr#24313](#), Patrick Donnelly)
- cephfs: ceph\_volume\_client: allow atomic update of RADOS objects ([issue#24173](#), [pr#24084](#), Rishabh Dave)
- cephfs: ceph\_volume\_client: delay required after adding data pool to MDSMap ([issue#25141](#), [pr#23726](#), Patrick Donnelly)
- cephfs: ceph\_volume\_client: py3 compatible ([issue#17230](#), [pr#24083](#), Rishabh Dave, Patrick Donnelly)
- cephfs: cephfs-data-scan: print the max used ino ([issue#26925](#), [pr#23881](#), "Yan, Zheng")
- cephfs: cephfs-journal-tool: wrong layout info used ([issue#24644](#), [pr#24033](#), Gu Zhongyan)
- cephfs: client: check for unmounted condition before printing debug output ([issue#25213](#), [pr#23617](#), Jeff Layton)
- cephfs: client: drop null child dentries before try pruning inode's alias ([issue#22293](#), [pr#24119](#), "Yan, Zheng")
- cephfs: client: fix choose\_target\_mds for requests that do name lookup ([issue#26860](#), [pr#23793](#), "Yan, Zheng")
- cephfs: client: retry remount on dcache invalidation failure ([issue#27657](#), [pr#24303](#), Venky Shankar)
- cephfs: client: statfs inode count odd ([issue#24849](#), [pr#24376](#), Rishabh Dave)
- cephfs: client: two ceph-fuse clients, one can not list out files created by another ([issue#27051](#), [pr#24282](#), Peng Xie)
- cephfs: client: update ctime when modifying file content ([issue#35945](#), [pr#24323](#), "Yan, Zheng")
- common: get real hostname from container/pod environment ([pr#23915](#), Sage Weil)
- core: PGPool::update optimizations ([pr#23969](#), Zac Medico)

- core: ceph-disk: compatibility fix for python 3 ([issue#35906](#), [pr#24347](#), Tim Serong)
- core: discover\_all\_missing() not always called during activating ([issue#22837](#), [pr#23817](#), Sage Weil, David Zafman)
- core: kv/KeyValueDB: return const char\* from MergeOperator::name() ([issue#26875](#), [pr#23566](#), Sage Weil)
- core: librados application's symbol could conflict with the libceph-common ([issue#25154](#), [pr#23483](#), Kefu Chai)
- core: mgr/MgrClient: guard send\_pgstats() with lock ([issue#23370](#), [pr#23791](#), Kefu Chai)
- core: mgr/balancer: deepcopy best plan - otherwise we get latest ([issue#27000](#), [pr#23740](#), Stefan Priebe)
- core: mgrc: enable disabling stats via mgr\_stats\_threshold ([issue#25197](#), [pr#23461](#), John Spray)
- core: mon/OSDMonitor: invalidate max\_failed\_since on cancel\_report ([issue#35860](#), [pr#24257](#), xie xingguo)
- core: object errors found in be\_select\_auth\_object() aren't logged the same ([issue#25108](#), [pr#23871](#), David Zafman)
- core: os/bluestore: bluestore\_buffer\_hit\_bytes perf counter doesn't reset ([pr#23773](#), Igor Fedotov)
- core: os/bluestore: cache autotuning and memory limit ([pr#24065](#), Mark Nelson)
- core: osd,mon: increase mon\_max\_pg\_per\_osd to 250 ([issue#25112](#), [pr#23862](#), Neha Ojha)
- core: osd/PG: avoid choose\_acting picking want with > pool size items ([issue#35924](#), [pr#24299](#), Sage Weil)
- core: osdc/Objecter: fix split vs reconnect race ([issue#22544](#), [pr#24188](#), Sage Weil)
- core: rados python bindings use prval from stack ([issue#25175](#), [pr#23864](#), Sage Weil)
- doc: Fix broken urls ([issue#25185](#), [pr#23621](#), Jos Collin)
- doc: remove deprecated 'scrubq' from ceph(8) ([issue#35813](#), [pr#24211](#), Ruben Kerkhof)
- doc: rgw: ldap-auth: fixed option name 'rgw\_ldap\_searchfilter' ([issue#23081](#), [pr#23761](#), Konstantin Shalygin)

- mds: MDBalancer::try\_rebalance() may stop prematurely ([issue#26973](#), [pr#23884](#), "Yan, Zheng")
- mds: allows client to create .. and . dirents ([issue#25113](#), [pr#24329](#), Venky Shankar)
- mds: avoid using g\_conf->get\_val<...>(...) in hot path ([issue#24820](#), [pr#23408](#), "Yan, Zheng")
- mds: calculate load by checking self CPU usage ([issue#26834](#), [pr#23505](#), "Yan, Zheng")
- mds: configurable timeout for client eviction ([issue#25188](#), [pr#24086](#), Patrick Donnelly, Venky Shankar)
- mds: crash when dumping ops in flight ([issue#26894](#), [pr#23677](#), "Yan, Zheng")
- mds: curate priority of perf counters sent to mgr ([issue#22097](#), [issue#24004](#), [pr#24089](#), Guan yunfei, Venky Shankar)
- mds: explain delayed client\_request due to subtree migration ([issue#24840](#), [pr#23678](#), Yan, Zheng, "Yan, Zheng")
- mds: health warning for slow metadata IO ([issue#24879](#), [pr#24171](#), "Yan, Zheng")
- mds: internal op missing events time 'throttled', 'all\_read', 'dispatched' ([issue#36114](#), [pr#24410](#), Yanhu Cao)
- mds: mds got laggy because of MDSBeacon stuck in mqueue ([issue#23519](#), [pr#23556](#), "Yan, Zheng")
- mds: optimize the way how max export size is enforced ([issue#25131](#), [pr#23789](#), "Yan, Zheng")
- mds: prevent MDSRank::evict\_client from blocking finisher thread ([issue#35720](#), [pr#23946](#), "Yan, Zheng")
- mds: print is\_laggy message once ([issue#35250](#), [pr#24138](#), Patrick Donnelly)
- mds: rctime may go back ([issue#35916](#), [pr#24378](#), "Yan, Zheng")
- mds: reset heartbeat map at potential time-consuming places ([issue#26858](#), [pr#23507](#), Yan, Zheng, "Yan, Zheng")
- mds: runs out of file descriptors after several respawns ([issue#35850](#), [pr#24310](#), Patrick Donnelly)
- mds: track average session uptime ([issue#25013](#), [pr#24421](#), Patrick Donnelly, Venky Shankar)
- mds: use monotonic clock for beacon message timekeeping ([issue#26959](#), [pr#24311](#), Patrick Donnelly)

- mgr: Sync the prometheus module ([pr#23216](#), Boris Ranto)
- mon: Automatically set expected\_num\_objects for new pools with >=100 PGs per OSD ([issue#24687](#), [pr#24395](#), Douglas Fuller)
- msg: “challenging authorizer” messages appear at debug\_ms=0 ([issue#35251](#), [pr#23943](#), Patrick Donnelly)
- msg: async: clean up local buffers on dispatch ([issue#35987](#), [pr#24387](#), Greg Farnum)
- msg: ceph\_abort() when there are enough accepter errors in msg server ([issue#23649](#), [pr#24419](#), penglaiyxy@gmail.com)
- osd: EC: slow/hung ops in multimds suite test ([issue#23769](#), [pr#24393](#), Sage Weil)
- osd: ECBackend: don't get result code of subchunk-read overwritten ([issue#21769](#), [pr#24342](#), songweibin)
- osd: Limit pg log length during recovery/backfill so that we don't run out of memory ([issue#21416](#), [pr#23211](#), Neha Ojha, xie xingguo)
- osd: OSDMap: fix apply upmap segfault ([issue#22056](#), [pr#23579](#), Brad Hubbard)
- osd: PG: add custom\_reaction Backfilled and release reservations after bac... ([issue#23614](#), [pr#23493](#), Neha Ojha)
- osd: PrimaryLogPG: fix potential pg-log overtrimming ([pr#24308](#), xie xingguo)
- osd: backport ‘bench’ and stdout changes ([issue#24022](#), [pr#23680](#), Коренберг Марк, John Spray, Kefu Chai)
- osd: read object attrs failed at EC recovery ([issue#24406](#), [pr#24327](#), xiaofei cui)
- osd: scrub livelock ([issue#26890](#), [pr#24396](#), Sage Weil)
- qa/suites/rados/upgrade/jewel-x-singleton: exclude python3-rados, python3-cephfs ([pr#24479](#), Neha Ojha)
- rbd: [rbd-mirror] failed assertion when updating mirror status ([issue#36084](#), [pr#24320](#), Jason Dillaman)
- rbd: fix error import when the input is a pipe ([issue#34536](#), [pr#24003](#), songweibin)
- rbd: librbd: blacklisted client might not notice it lost the lock ([issue#34534](#), [pr#24405](#), Song Shun, Mykola Golub, Jason Dillaman)
- rbd: librbd: discard should wait for in-flight cache writeback to complete ([issue#23548](#), [pr#23594](#), Jason Dillaman)
- rbd: librbd: ensure exclusive lock acquired when removing sync point snaps...

- ([issue#24898](#), [pr#24123](#), Mykola Golub, Jason Dillaman)
- rbd: librbd: fix refuse to release lock when cookie is the same at rewatch ([issue#27986](#), [pr#23758](#), Song Shun)
- rbd: librbd: fixed assert when flattening clone with zero overlap ([issue#35702](#), [pr#24285](#), Jason Dillaman)
- rbd: librbd: image create request should validate data pool for self-managed snapshot support ([issue#24675](#), [pr#24390](#), Mykola Golub)
- rbd: librbd: journaling unable request can not be sent to remote lock owner ([issue#26939](#), [pr#24100](#), Mykola Golub)
- rbd: librbd: object map improperly flagged as invalidated ([issue#24516](#), [pr#24415](#), Jason Dillaman)
- rbd: librbd: potential race on image create request complete ([issue#24910](#), [pr#23892](#), Mykola Golub)
- rgw: 'radosgw-admin sync error trim' only trims partially ([issue#24873](#), [pr#24054](#), Casey Bodley)
- rgw: Fix log level of gc\_iterate\_entries ([issue#23801](#), [pr#23665](#), iliul)
- rgw: Limit the number of lifecycle rules on one bucket ([issue#24572](#), [pr#23522](#), Zhang Shaowen)
- rgw: The delete markers generated by object expiration should have owner ([issue#24568](#), [pr#23545](#), Zhang Shaowen)
- rgw: abort\_bucket\_multiparts() ignores individual NoSuchUpload errors ([issue#35986](#), [pr#24389](#), Casey Bodley)
- rgw: change default rgw\_thread\_pool\_size to 512 ([issue#25214](#), [issue#24544](#), [pr#24034](#), Douglas Fuller, Casey Bodley)
- rgw: cls/rgw: don't assert in decode\_list\_index\_key() ([issue#24117](#), [pr#24391](#), Yehuda Sadeh)
- rgw: cls/rgw: ready rgw\_usage\_log\_entry for extraction via ceph-dencoder ([issue#34537](#), [pr#23974](#), Vaibhav Bhembre)
- rgw: fix chunked-encoding for chunks >1MiB ([issue#35990](#), [pr#24361](#), Robin H. Johnson)
- rgw: fix deadlock on RGWIndexCompletionManager::stop ([issue#26949](#), [pr#24069](#), Yao Zongyou)
- rgw: incremental data sync uses truncated flag to detect end of listing ([issue#26952](#), [pr#24242](#), Casey Bodley)

- rgw: multisite: data sync error repo processing does not back off on empty ([issue#26938](#), [pr#24318](#), Casey Bodley)
- rgw: multisite: intermittent failures in test\_bucket\_sync\_disable\_enable ([issue#26895](#), [pr#24316](#), Casey Bodley)
- rgw: multisite: intermittent test\_bucket\_index\_log\_trim failures ([issue#36034](#), [pr#24398](#), Casey Bodley)
- rgw: multisite: object metadata operations are skipped by sync ([issue#24367](#), [pr#24056](#), Casey Bodley)
- rgw: multisite: object name should be urlencoded when we put it into ES ([issue#23216](#), [pr#24424](#), Chang Liu)
- rgw: multisite: out of order updates to sync status markers ([issue#35539](#), [pr#24317](#), Yehuda Sadeh)
- rgw: multisite: segfault on shutdown/realm reload ([issue#35543](#), [pr#24231](#), Casey Bodley)
- rgw: multisite: update index segfault on shutdown/realm reload ([issue#35905](#), [pr#24397](#), Tianshan Qu)
- rgw: raise debug level on redundant data sync error messages ([issue#35830](#), [issue#36037](#), [pr#24135](#), Casey Bodley, Matt Benjamin)
- rgw: raise default rgw\_curl\_low\_speed\_time to 300 seconds ([issue#27989](#), [pr#24046](#), Casey Bodley)
- rgw: resharding produces invalid values of bucket stats ([issue#36290](#), [pr#24527](#), Abhishek Lekshmanan)
- rgw: return x-amz-version-id: null when delete obj in versioning suspended bucket ([issue#35814](#), [pr#24190](#), yuliyang)
- rgw: rgw\_file: deep stat handling ([issue#24915](#), [pr#23499](#), Matt Benjamin)
- tests: Excluded 'python34-cephfs' from the install tasks ([pr#24650](#), Yuri Weinstein)
- tests: Use pids instead of jobspecs which were wrong ([issue#27056](#), [pr#23901](#), David Zafman)
- tests: cephfs: multifs requires 4 mds but gets only 2 ([issue#24899](#), [pr#24328](#), Patrick Donnelly)
- tests: cls\_rgw test is only run in rados suite: add it to rgw suite as well ([issue#24815](#), [pr#24070](#), Casey Bodley, Sage Weil)
- tests: librbd: not valid to have different parents between image snapshots

([issue#36097](#), [pr#24245](#), Jason Dillaman)

- tests: move mds/client config to qa from teuthology ceph.conf.template ([issue#26900](#), [issue#24839](#), [pr#23877](#), Patrick Donnelly)
- tests: qa/tasks: s3a fix mirror ([pr#24039](#), Vasu Kulkarni)
- tests: qa/workunits: replace ‘realpath’ with ‘readlink -f’ in fsstress.sh ([issue#27211](#), [issue#36409](#), [pr#24620](#), Ilya Dryomov, Jason Dillaman)
- tests: qa: add .qa helper link ([pr#24134](#), Patrick Donnelly)
- tests: qa: added v12.2.8 to the mix ([issue#35541](#), [pr#23913](#), Yuri Weinstein)
- tests: remove knfs qa suite from future releases ([issue#36075](#), [pr#24268](#), Yuri Weinstein)
- tools: ceph-objectstore-tool: Allow target level as first positional parameter ([issue#35846](#), [pr#24115](#), David Zafman)

## v12.2.8 Luminous

---

This is the eighth bug fix release of the Luminous v12.2.x long term stable release series. This release contains several bugfixes across all the components and we recommend all users upgrade.

## Upgrade Notes from previous luminous releases

---

When upgrading from v12.2.5 or v12.2.6 please note that upgrade caveats from 12.2.5 will apply to any \_newer\_ luminous version including 12.2.8. Please read the notes at [luminous-12-2-5-upgrades](#) .

For the cluster that installed the broken 12.2.6 release, 12.2.7 fixed the regression and introduced a workaround option osd distrust data digest = true, but 12.2.7 clusters still generated health warnings like

```
[ERR] 11.288 shard 207: soid 11:1155c332:::rbd_data.207dce238e1f29.0000000000000527:head data_digest
1. 0xc8997a5b != data_digest 0x2ca15853
```

12.2.8 improves the deep scrub code to automatically repair these inconsistencies. Once the entire cluster has been upgraded and then fully deep scrubbed, and all such inconsistencies are resolved, it will be safe to disable the osd distrust data digest = true workaround option.

## Notable Changes

---

- OSD

- Scrub repair is enhanced to handle data digest mismatch info on replicas as long as all replicas' digests match each other.
- RGW
  - Options rgw curl low speed limit and rgw curl low speed time are added to control the lower speed limits and times below which the requests are considered too slow to be aborted and can help mitigate data sync getting blocked during network issues
  - Option rgw s3 auth order configurable added which takes a comma separated list of order to try for s3 authentication when external engines are involved.

## Changelog

---

- bluestore: set correctly shard for existed Collection ([issue#24761](#), [pr#22860](#), Jianpeng Ma)
- build/ops: Boost system library is no longer required to compile and link example librados program ([issue#25054](#), [pr#23202](#), Nathan Cutler)
- build/ops: Bring back diff -y for non-FreeBSD ([issue#24396](#), [issue#21664](#), [pr#22848](#), Sage Weil, David Zafman)
- build/ops: install-deps.sh fails on newest openSUSE Leap ([issue#25064](#), [pr#23179](#), Kyr Shatskyy)
- build/ops: Mimic build fails with -DWITH\_RADOSGW=0 ([issue#24437](#), [pr#22864](#), Dan Mick)
- build/ops: order rbdmap.service before remote-fs-pre.target ([issue#24713](#), [pr#22844](#), Ilya Dryomov)
- build/ops: rpm: silence osd block chown ([issue#25152](#), [pr#23313](#), Dan van der Ster)
- cephfs-journal-tool: Fix purging when importing an zero-length journal ([issue#24239](#), [pr#22980](#), yupeng chen, zhongyan gu)
- cephfs: MDSMonitor: uncommitted state exposed to clients/mdss ([issue#23768](#), [pr#23013](#), Patrick Donnelly)
- ceph-fuse mount failed because no mds ([issue#22205](#), [pr#22895](#), liyan)
- ceph-volume add a \_\_release\_\_ string, to help version-conditional calls ([issue#25170](#), [pr#23331](#), Alfredo Deza)
- ceph-volume: adds test for ceph-volume lvm list /dev/sda ([issue#24784](#), [issue#24957](#), [pr#23350](#), Andrew Schoen)

- ceph-volume: do not use stdin in luminous ([issue#25173](#), [issue#23260](#), [pr#23367](#), Alfredo Deza)
- ceph-volume enable the ceph-osd during lvm activation ([issue#24152](#), [pr#23394](#), Dan van der Ster, Alfredo Deza)
- ceph-volume expand on the LVM API to create multiple LVs at different sizes ([issue#24020](#), [pr#23395](#), Alfredo Deza)
- ceph-volume lvm.activate conditional mon-config on prime-osd-dir ([issue#25216](#), [pr#23397](#), Alfredo Deza)
- ceph-volume lvm.batch remove non-existent sys\_api property ([issue#34310](#), [pr#23811](#), Alfredo Deza)
- ceph-volume lvm.listing only include devices if they exist ([issue#24952](#), [pr#23150](#), Alfredo Deza)
- ceph-volume: process.call with stdin in Python 3 fix ([issue#24993](#), [pr#23238](#), Alfredo Deza)
- ceph-volume: PVolumes.get() should return one PV when using name or uuid ([issue#24784](#), [pr#23329](#), Andrew Schoen)
- ceph-volume: refuse to zap mapper devices ([issue#24504](#), [pr#23374](#), Andrew Schoen)
- ceph-volume: tests.functional inherit SSH\_ARGS from ansible ([issue#34311](#), [pr#23813](#), Alfredo Deza)
- ceph-volume tests/functional run lvm list after OSD provisioning ([issue#24961](#), [pr#23147](#), Alfredo Deza)
- ceph-volume: unmount lvs correctly before zapping ([issue#24796](#), [pr#23128](#), Andrew Schoen)
- ceph-volume: update batch documentation to explain filestore strategies ([issue#34309](#), [pr#23825](#), Alfredo Deza)
- change default filestore\_merge\_threshold to -10 ([issue#24686](#), [pr#22814](#), Douglas Fuller)
- client: add inst to asok status output ([issue#24724](#), [pr#23107](#), Patrick Donnelly)
- client: fixup parallel calls to ceph\_ll\_lookup\_inode() in NFS FASL ([issue#22683](#), [pr#23012](#), huanwen ren)
- client: increase verbosity level for log messages in helper methods ([issue#21014](#), [pr#23014](#), Rishabh Dave)
- client: update inode fields according to issued caps ([issue#24269](#), [pr#22783](#), "Yan, Zheng")

- common: Abort in OSDMap::decode() during qa/standalone/erasure-code/test-erasure-eio.sh ([issue#23492](#), [pr#23025](#), Sage Weil)
- common/DecayCounter: set last\_decay to current time when decoding decay counter ([issue#24440](#), [pr#22779](#), Zhi Zhang)
- doc: ceph-bluestore-tool manpage not getting rendered correctly ([issue#24800](#), [pr#23177](#), Nathan Cutler)
- filestore: add pgid in filestore pg dir split log message ([issue#24878](#), [pr#23454](#), Vikhyat Umrao)
- let “ceph status” use base 10 when printing numbers not sizes ([issue#22095](#), [pr#22680](#), Jan Fajerski, Kefu Chai)
- librados: fix buffer overflow for aio\_exec python binding ([issue#23964](#), [pr#22708](#), Aleksei Gutikov)
- librbd: force ‘invalid object map’ flag on-disk update ([issue#24434](#), [pr#22753](#), Mykola Golub)
- librbd: utilize the journal disabled policy when removing images ([issue#23512](#), [pr#23595](#), Jason Dillaman)
- mds: don’t report slow request for blocked filelock request ([issue#22428](#), [pr#22782](#), “Yan, Zheng”)
- mds: dump recent events on respawn ([issue#24853](#), [pr#23213](#), Patrick Donnelly)
- mds: handle discontinuous mdsmap ([issue#24856](#), [pr#23169](#), “Yan, Zheng”)
- mds: increase debug level for dropped client cap msg ([issue#24855](#), [pr#23214](#), Patrick Donnelly)
- mds: low wrlock efficiency due to dirfrags traversal ([issue#24467](#), [pr#22885](#), Xuehan Xu)
- mds: print mdsmap processed at low debug level ([issue#24852](#), [pr#23212](#), Patrick Donnelly)
- mds: scrub doesn’t always return JSON results ([issue#23958](#), [pr#23222](#), Venky Shankar)
- mds: unset deleted vars in shutdown\_pass ([issue#23766](#), [pr#23015](#), Patrick Donnelly)
- mgr: add units to performance counters ([issue#22747](#), [pr#23266](#), Ernesto Puerta, Rubab Syed)
- mgr: ceph osd safe-to-destroy crashes the mgr ([issue#23249](#), [pr#22806](#), Sage Weil)
- mgr/MgrClient: Protect daemon\_health\_metrics ([issue#23352](#), [pr#23459](#), Kjetil

Joergensen, Brad Hubbard)

- mon: Add option to view IP addresses of clients in output of 'ceph features' ([issue#21315](#), [pr#22773](#), Paul Emmerich)
- mon/HealthMonitor: do not send MMonHealthChecks to pre-luminous mon ([issue#24481](#), [pr#22655](#), Sage Weil)
- os/bluestore: fix flush\_commit locking ([issue#21480](#), [pr#22904](#), Sage Weil)
- os/bluestore: fix incomplete faulty range marking when doing compression ([issue#21480](#), [pr#22909](#), Igor Fedotov)
- os/bluestore: fix races on SharedBlob::coll in ~SharedBlob ([issue#24859](#), [pr#23064](#), Radoslaw Zarzynski)
- osdc: Fix the wrong BufferHead offset ([issue#24484](#), [pr#22865](#), dongdong tao)
- osd: do\_sparse\_read(): Verify checksum earlier so we will try to repair and missed backport ([issue#24875](#), [pr#23379](#), xie xingguo, David Zafman)
- osd: eternal stuck PG in 'unfound\_recovery' ([issue#24373](#), [pr#22546](#), Sage Weil)
- osd: may get empty info at recovery ([issue#24588](#), [pr#22862](#), Sage Weil)
- osd/OSDMap: CRUSH\_TUNABLES5 added in jewel, not kraken ([issue#25057](#), [pr#23227](#), Sage Weil)
- osd/Session: fix invalid iterator dereference in Sessoin::have\_backoff() ([issue#24486](#), [pr#22729](#), Sage Weil)
- pjd: cd: too many arguments ([issue#24307](#), [pr#22883](#), Neha Ojha)
- PurgeQueue sometimes ignores Journaler errors ([issue#24533](#), [pr#22811](#), John Spray)
- pybind: pybind/mgr/mgr\_module: make rados handle available to all modules ([issue#24788](#), [issue#25102](#), [pr#23235](#), Ernesto Puerta, Sage Weil)
- pybind: Python bindings use iteritems method which is not Python 3 compatible ([issue#24779](#), [pr#22918](#), Nathan Cutler, Kefu Chai)
- pybind: rados.pyx: make all exceptions accept keyword arguments ([issue#24033](#), [pr#22979](#), Rishabh Dave)
- rbd: fix issues in IEC unit handling ([issue#26927](#), [issue#26928](#), [pr#23776](#), Jason Dillaman)
- repeated eviction of idle client until some IO happens ([issue#24052](#), [pr#22780](#), "Yan, Zheng")
- rgw: add curl\_low\_speed\_limit and curl\_low\_speed\_time config to avoid the thread hangs in data sync ([issue#25019](#), [pr#23144](#), Mark Kogan, Zhang Shaowen)

- rgw: add unit test for cls bi list command ([issue#24483](#), [pr#22846](#), Orit Wasserman, Xinying Song)
- rgw: do not ignore EEXIST in RGWPutObj::execute ([issue#22790](#), [pr#23207](#), Matt Benjamin)
- rgw: fail to recover index from crash luminous backport ([issue#24640](#), [issue#24280](#), [pr#23130](#), Tianshan Qu)
- rgw: fix gc may cause a large number of read traffic ([issue#24767](#), [pr#22984](#), Xin Liao)
- rgw: fix the bug of radogw-admin zonegroup set requires realm ([issue#21583](#), [pr#22767](#), lvshanchun)
- rgw: have a configurable authentication order ([issue#23089](#), [pr#23501](#), Abhishek Lekshmanan)
- rgw: index complete miss zones\_trace set ([issue#24590](#), [pr#22820](#), Tianshan Qu)
- rgw: Invalid Access-Control-Request may bypass validate\_cors\_rule\_method ([issue#24223](#), [pr#22934](#), Jeegn Chen)
- rgw: meta and data notify thread miss stop cr manager ([issue#24589](#), [pr#22822](#), Tianshan Qu)
- rgw-multisite: endless loop in RGWBucketShardIncrementalSyncCR ([issue#24603](#), [pr#22817](#), cfanz)
- rgw performance regression for luminous 12.2.4 ([issue#23379](#), [pr#22930](#), Mark Kogan)
- rgw: radogw-admin reshards status command should print text for reshards ([issue#23257](#), [pr#23019](#), Orit Wasserman)
- rgw: "radosgw-admin objects expire" always returns ok even if the probe fails ([issue#24592](#), [pr#23000](#), Zhang Shaowen)
- rgw: require -yes-i-really-mean-it to run radosgw-admin orphans find ([issue#24146](#), [pr#22985](#), Matt Benjamin)
- rgw: REST admin metadata API paging failure bucket & bucket.instance: InvalidArgument ([issue#23099](#), [pr#22932](#), Matt Benjamin)
- rgw: set cr state if aio\_read err return in RGWCloneMetaLogCoroutine ([issue#24566](#), [pr#22942](#), Tianshan Qu)
- spdk: fix ceph-osd crash when activate SPDK ([issue#24371](#), [pr#22686](#), tone-zhang)
- tools/ceph-objectstore-tool: split filestore directories offline to target hash level ([issue#21366](#), [pr#23418](#), Zhi Zhang)

## v12.2.7 Luminous

This is the seventh bugfix release of Luminous v12.2.x long term stable release series. This release contains several fixes for regressions in the v12.2.6 and v12.2.5 releases. We recommend that all users upgrade.

### note

The v12.2.6 release has serious known regressions. If you installed this release, please see the upgrade procedure below.

### note

The v12.2.5 release has a potential data corruption issue with erasure coded pools. If you ran v12.2.5 with erasure coding, please see below.

## Upgrading from v12.2.6

v12.2.6 included an incomplete backport of an optimization for BlueStore OSDs that avoids maintaining both the per-object checksum and the internal BlueStore checksum. Due to the accidental omission of a critical follow-on patch, v12.2.6 corrupts (fails to update) the stored per-object checksum value for some objects. This can result in an EIO error when trying to read those objects.

1. If your cluster uses FileStore only, no special action is required. This problem only affects clusters with BlueStore.
2. If your cluster has only BlueStore OSDs (no FileStore), then you should enable the following OSD option:

```
1. osd skip data digest = true
```

This will avoid setting and start ignoring the full-object digests whenever the primary for a PG is BlueStore.

3. If you have a mix of BlueStore and FileStore OSDs, then you should enable the following OSD option:

```
1. osd distrust data digest = true
```

This will avoid setting and start ignoring the full-object digests in all cases. This weakens the data integrity checks for FileStore (although those checks were always only opportunistic).

If your cluster includes BlueStore OSDs and was affected, deep scrubs will generate errors about mismatched CRCs for affected objects. Currently the repair operation does not know how to correct them (since all replicas do not match the expected checksum it

does not know how to proceed). These warnings are harmless in the sense that IO is not affected and the replicas are all still in sync. The number of affected objects is likely to drop (possibly to zero) on their own over time as those objects are modified. We expect to include a scrub improvement in v12.2.8 to clean up any remaining objects.

Additionally, see the notes below, which apply to both v12.2.5 and v12.2.6.

## Upgrading from v12.2.5 or v12.2.6

---

If you used v12.2.5 or v12.2.6 in combination with erasure coded pools, there is a small risk of corruption under certain workloads. Specifically, when:

- An erasure coded pool is in use
- The pool is busy with successful writes
- The pool is also busy with updates that result in an error result to the librados user. RGW garbage collection is the most common example of this (it sends delete operations on objects that don't always exist.)
- Some OSDs are reasonably busy. One known example of such load is FileStore splitting, although in principle any load on the cluster could also trigger the behavior.
- One or more OSDs restarts.

This combination can trigger an OSD crash and possibly leave PGs in a state where they fail to peer.

Notably, upgrading a cluster involves OSD restarts and as such may increase the risk of encountering this bug. For this reason, for clusters with erasure coded pools, we recommend the following upgrade procedure to minimize risk:

1. Install the v12.2.7 packages.
2. Temporarily quiesce IO to cluster:

```
1. ceph osd pause
```

3. Restart all OSDs and wait for all PGs to become active.
4. Resume IO:

```
1. ceph osd unpause
```

This will cause an availability outage for the duration of the OSD restarts. If this is unacceptable, an *more risky* alternative is to disable RGW garbage collection (the

primary known cause of these rados operations) for the duration of the upgrade:

1. #. Set ``rgw\_enable\_gc\_threads = false`` in ceph.conf
2. #. Restart all radosgw daemons
3. #. Upgrade and restart all OSDs
4. #. Remove ``rgw\_enable\_gc\_threads = false`` from ceph.conf
5. #. Restart all radosgw daemons

## Upgrading from other versions

---

If your cluster did not run v12.2.5 or v12.2.6 then none of the above issues apply to you and you should upgrade normally.

## Notable Changes

---

- mon/AuthMonitor: improve error message ([issue#21765](#), [pr#22963](#), Douglas Fuller)
- osd/PG: do not blindly roll forward to log.head ([issue#24597](#), [pr#22976](#), Sage Weil)
- osd/PrimaryLogPG: rebuild attrs from clients ([issue#24768](#), [pr#22962](#), Sage Weil)
- osd: work around data digest problems in 12.2.6 (version 2) ([issue#24922](#), [pr#23055](#), Sage Weil)
- rgw: objects in cache never refresh after rgw\_cache\_expiry\_interval ([issue#24346](#), [pr#22369](#), Casey Bodley, Matt Benjamin)

## v12.2.6 Luminous

### note

This is a broken release with serious known regressions. Do not install it.

This is the sixth bugfix release of Luminous v12.2.x long term stable release series. This release contains a range of bug fixes across all components of Ceph and a few security fixes.

## Notable Changes

- *Auth:*
  - In 12.2.4 and earlier releases, keyring caps were not checked for validity, so the caps string could be anything. As of 12.2.6, caps strings are validated and providing a keyring with an invalid caps string to, e.g., “ceph auth add” will result in an error.
  - CVE 2018-1128: auth: cephx authorizer subject to replay attack ([issue#24836](#), Sage Weil)
  - CVE 2018-1129: auth: cephx signature check is weak ([issue#24837](#), Sage Weil)
  - CVE 2018-10861: mon: auth checks not correct for pool ops ([issue#24838](#), Jason Dillaman)
- The config-key interface can store arbitrary binary blobs but JSON can only express printable strings. If binary blobs are present, the ‘ceph config-key dump’ command will show them as something like `<<< binary blob of length N >>>`.

## Other Notable Changes

- build/ops: build-integration-branch script ([issue#24003](#), [pr#21919](#), Nathan Cutler, Kefu Chai, Sage Weil)
- cephfs-journal-tool: wait prezero ops before destroying journal ([issue#20549](#), [pr#21874](#), “Yan, Zheng”)
- cephfs: MDSMonitor: cleanup and protect fsmap access ([issue#23762](#), [pr#21732](#), Patrick Donnelly)
- cephfs: MDSMonitor: crash after assigning standby-replay daemon in multifs setup ([issue#23762](#), [issue#23658](#), [pr#22603](#), “Yan, Zheng”)
- cephfs: MDSMonitor: fix mds health printed in bad format ([issue#23582](#), [pr#21447](#), Patrick Donnelly)

- cephfs: MDSMonitor: initialize new Filesystem epoch from pending ([issue#23764](#), [pr#21512](#), Patrick Donnelly)
- ceph-fuse: missing dentries in readdir result ([issue#23894](#), [pr#22119](#), "Yan, Zheng")
- ceph-fuse: return proper exit code ([issue#23665](#), [pr#21495](#), Patrick Donnelly)
- ceph-fuse: trim ceph-fuse -V output ([issue#23248](#), [pr#21600](#), Jos Collin)
- ceph\_test\_rados\_api\_aio: fix race with full pool and osdmap ([issue#23917](#), [issue#23876](#), [pr#21778](#), Sage Weil)
- ceph-volume: error on commands that need ceph.conf to operate ([issue#23941](#), [pr#22746](#), Andrew Schoen)
- ceph-volume: failed ceph-osd -mkfs command doesn't halt the OSD creation process ([issue#23874](#), [pr#21746](#), Alfredo Deza)
- client: add ceph\_ll\_sync\_inode ([issue#23291](#), [pr#21109](#), Jeff Layton)
- client: add client option descriptions ([issue#22933](#), [pr#21589](#), Patrick Donnelly)
- client: anchor dentries for trimming to make cap traversal safe ([issue#24137](#), [pr#22201](#), Patrick Donnelly)
- client: avoid freeing inode when it contains TX buffer head ([issue#23837](#), [pr#22168](#), Guan yunfei, "Yan, Zheng", Jason Dillaman)
- client: dirty caps may never get the chance to flush ([issue#22546](#), [pr#21278](#), dongdong tao)
- client: fix issue of revoking non-auth caps ([issue#24172](#), [pr#22221](#), "Yan, Zheng")
- client: fix request send\_to\_auth was never really used ([issue#23541](#), [pr#21354](#), Zhi Zhang)
- client: Fix the gid\_count check ([issue#23652](#), [pr#21596](#), Jos Collin)
- client: flush the mdlog in \_fsync before waiting on unstable reqs ([issue#23714](#), [pr#21542](#), Jeff Layton)
- client: hangs on umount if it had an MDS session evicted ([issue#10915](#), [pr#22018](#), Rishabh Dave)
- client: void sending mds request while holding cap reference ([issue#24369](#), [pr#22354](#), "Yan, Zheng")
- cmake: fix the cepfs java binding build on Bionic ([issue#23458](#), [issue#24012](#), [pr#21872](#), Kefu Chai, Shengjing Zhu)
- cmake/modules/BuildRocksDB.cmake: enable compressions for rocksdb ([issue#24025](#),

pr#22215, Kefu Chai)

- common: ARMv8 feature detection broken, leading to illegal instruction crashes ([issue#23464](#), [pr#22567](#), Adam Kupczyk)
- common: fix BoundedKeyCounter const\_pointer\_iterator ([issue#22139](#), [pr#21083](#), Casey Bodley)
- common: fix typo in rados bench write JSON output ([issue#24199](#), [pr#22391](#), Sandor Zeestraten)
- common: partially revert 95fc248 to make get\_process\_name work ([issue#24123](#), [pr#22290](#), Mykola Golub)
- core: Deleting a pool with active notify linger ops can result in seg fault ([issue#23966](#), [pr#22143](#), Kefu Chai, Jason Dillaman)
- core: mon/MgrMonitor: change 'unresponsive' message to info level ([issue#24222](#), [pr#22331](#), Sage Weil)
- core: Wip scrub omap ([issue#24366](#), [pr#22375](#), xie xingguo, David Zafman)
- crush: fix device\_class\_clone for unpopulated/empty weight-sets ([issue#23386](#), [pr#22381](#), Sage Weil)
- crush, osd: handle multiple parents properly when applying pg upmaps ([issue#23921](#), [pr#22115](#), xiexingguo)
- doc: Fix -d description in ceph-fuse ([issue#23214](#), [pr#21616](#), Jos Collin)
- doc: Update ceph-fuse doc ([issue#23084](#), [pr#21603](#), Jos Collin)
- fuse: wire up fuse\_ll\_access ([issue#23509](#), [pr#21475](#), Jeff Layton)
- kceph: umount on evicted client blocks forever ([issue#24053](#), [issue#24054](#), [pr#22208](#), Yan, Zheng, "Yan, Zheng")
- librbd: commit IO as safe when complete if writeback cache is disabled ([issue#23516](#), [pr#22370](#), Jason Dillaman)
- librbd: prevent watcher from unregistering with in-flight actions ([issue#23955](#), [pr#21938](#), Jason Dillaman)
- lvm: when osd creation fails log the exception ([issue#24456](#), [pr#22641](#), Andrew Schoen)
- mds: avoid calling rejoин\_gather\_finish() two times successively ([issue#24047](#), [pr#22171](#), "Yan, Zheng")
- mds: broadcast quota to relevant clients when quota is explicitly set ([issue#24133](#), [pr#22271](#), Zhi Zhang)

- mds: crash when failover ([issue#23518](#), [pr#21900](#), "Yan, Zheng")
- mds: don't discover inode/dirfrag when mds is in 'starting' state ([issue#23812](#), [pr#21990](#), "Yan, Zheng")
- mds: fix occasional dir rstat inconsistency between multi-MDSes ([issue#23538](#), [pr#21617](#), "Yan, Zheng", Zhi Zhang)
- mds: fix some memory leak ([issue#24289](#), [pr#22310](#), "Yan, Zheng")
- mds: fix unhealth heartbeat during rejoin ([issue#23530](#), [pr#21366](#), dongdong tao)
- mds: handle imported session race ([issue#24072](#), [issue#24087](#), [pr#21989](#), Patrick Donnelly)
- mds: include nfiles/nsubdirs of directory inode in MClientCaps ([issue#23855](#), [pr#22118](#), "Yan, Zheng")
- mds: kick rdlock if waiting for dirfragtreelock ([issue#23919](#), [pr#21901](#), Patrick Donnelly)
- mds: make rstat.rctime follow inodes' ctime ([issue#23380](#), [pr#21448](#), "Yan, Zheng")
- mds: mark damaged if sessions' preallocated inos don't match inotable ([issue#23452](#), [pr#21372](#), "Yan, Zheng")
- mds: mark new root inode dirty ([issue#23960](#), [pr#21922](#), Patrick Donnelly)
- mds: mds shutdown fixes and optimization ([issue#23602](#), [pr#21346](#), "Yan, Zheng")
- mds: misc load balancer fixes ([issue#21745](#), [pr#21412](#), "Yan, Zheng", Jianyu Li)
- mds: properly check auth subtree count in MDCache::shutdown\_pass() ([issue#23813](#), [pr#21844](#), "Yan, Zheng")
- mds: properly dirty sessions opened by journal replay ([issue#23625](#), [pr#21441](#), "Yan, Zheng")
- mds: properly trim log segments after scrub repairs something ([issue#23880](#), [pr#21840](#), "Yan, Zheng")
- mds: set could\_consume to false when no purge queue item actually exe... ([issue#24073](#), [pr#22176](#), Xuehan Xu)
- mds: trim log during shutdown to clean metadata ([issue#23923](#), [pr#21899](#), Patrick Donnelly)
- mds: underwater dentry check in CDir::\_omap\_fetched is racy ([issue#23032](#), [pr#21187](#), Yan, Zheng)
- mg\_read() call has wrong arguments ([issue#23596](#), [pr#21382](#), Nathan Cutler)

- mgr/influx: Only split string on first occurrence of dot (.) ([issue#23996](#), [pr#21965](#), Wido den Hollander)
- mgr: Module 'balancer' has failed: could not find bucket -14 ([issue#24167](#), [pr#22308](#), Sage Weil)
- mon: add 'ceph osd pool get erasure allow\_ec\_overwrites' command ([issue#23487](#), [pr#21378](#), Mykola Golub)
- mon: enable level\_compaction\_dynamic\_level\_bytes for rocksdb ([issue#24361](#), [pr#22360](#), Kefu Chai)
- mon: handle bad snapshot removal reqs gracefully ([issue#18746](#), [pr#21717](#), Paul Emmerich)
- mon: High MON cpu usage when cluster is changing ([issue#23713](#), [pr#21968](#), Sage Weil, Xiaoxi CHEN)
- mon/MDSMonitor: do not send redundant MDS health messages to cluster log ([issue#24308](#), [pr#22558](#), Sage Weil)
- msg/async/AsyncConnection: Fix FPE in process\_connection ([issue#23618](#), [pr#21376](#), Brad Hubbard)
- os/bluestore: alter the allow\_eio policy regarding kernel's error list ([issue#23333](#), [pr#21405](#), Radoslaw Zarzynski)
- os/bluestore/bluefs\_types: make block\_mask 64-bit ([issue#23840](#), [pr#21740](#), Sage Weil)
- os/bluestore: fix exceeding the max IO queue depth in KernelDevice ([issue#23246](#), [pr#21407](#), Radoslaw Zarzynski)
- os/bluestore: fix SharedBlobSet refcounting race ([issue#24319](#), [pr#22650](#), Sage Weil)
- os/bluestore: simplify and fix SharedBlob::put() ([issue#24211](#), [pr#22351](#), Sage Weil)
- osdc/Objecter: fix recursive locking in \_finish\_command ([issue#23940](#), [pr#21939](#), Sage Weil)
- osdc/Objecter: prevent double-invocation of linger op callback ([issue#23872](#), [pr#21752](#), Jason Dillaman)
- osd: do not crash on empty snapset ([issue#23851](#), [pr#21638](#), Mykola Golub, Igor Fedotov)
- osd: Don't evict even when preemption has restarted with smaller chunk ([issue#22881](#), [issue#23909](#), [issue#23646](#), [pr#22044](#), Sage Weil, fang yuxiang, Jianpeng Ma, kungf, xie xingguo, David Zafman)

- osd/ECBackend: only check required shards when finishing recovery reads ([issue#23195](#), [pr#21911](#), Josh Durgin, Kefu Chai)
- osd: increase default hard pg limit ([issue#24243](#), [pr#22592](#), Josh Durgin)
- osd/OSDMap: check against cluster topology changing before applying pg upmaps ([issue#23878](#), [pr#21818](#), xiexingguo)
- osd/PG: fix DeferRecovery vs AllReplicasRecovered race ([issue#23860](#), [pr#21964](#), Sage Weil)
- osd/PG: fix uninit read in Incomplete::react(AdvMap&) ([issue#23980](#), [pr#21993](#), Sage Weil)
- osd/PrimaryLogPG: avoid infinite loop when flush collides with write ... ([issue#23664](#), [pr#21764](#), Sage Weil)
- osd: publish osdmap to OSDService before starting wq threads ([issue#21977](#), [pr#21737](#), Sage Weil)
- osd: Warn about objects with too many omap entries ([issue#23784](#), [pr#21518](#), Brad Hubbard)
- qa: disable -Werror when compiling env\_librados\_test ([issue#23786](#), [pr#21655](#), Kefu Chai)
- qa: fix blacklisted check for test\_lifecycle ([issue#23975](#), [pr#21921](#), Patrick Donnelly)
- qa: remove racy/buggy test\_purge\_queue\_op\_rate ([issue#23829](#), [pr#21841](#), Patrick Donnelly)
- qa/suites/rbd/basic/msgr-failures: remove many.yaml ([issue#23789](#), [pr#22128](#), Sage Weil)
- qa: wait longer for osd to flush pg stats ([issue#24321](#), [pr#22296](#), Kefu Chai)
- qa/workunits/mon/test\_mon\_config\_key.py fails on master ([issue#23622](#), [pr#21368](#), Sage Weil)
- qa/workunits/rbd: adapt import\_export test to handle multiple units ([issue#24733](#), [pr#22911](#), Jason Dillaman)
- qa/workunits/rbd: potential race in mirror disconnect test ([issue#23938](#), [pr#21869](#), Mykola Golub)
- radosgw-admin sync status improvements ([issue#20473](#), [pr#21908](#), lvshanchun, Casey Bodley)
- rbd: improve 'import-diff' corrupt input error messages ([issue#18844](#), [issue#23038](#), [pr#21316](#), PCzhangPC, songweibin, Jason Dillaman)

- rbd-mirror: ensure remote demotion is replayed locally ([issue#24009](#), [pr#22142](#), Jason Dillaman)
- rbd-nbd can deadlock in logging thread ([issue#23143](#), [pr#21705](#), Sage Weil)
- rbd: python bindings fixes and improvements ([issue#23609](#), [pr#21725](#), Ricardo Dias)
- rbd: [rbd-mirror] asok hook for image replayer not re-registered after bootstrap ([issue#23888](#), [pr#21726](#), Jason Dillaman)
- rbd: [rbd-mirror] local tag predecessor mirror uuid is incorrectly replaced with remote ([issue#23876](#), [pr#21741](#), Jason Dillaman)
- rbd: [rbd-mirror] potential deadlock when running asok 'flush' command ([issue#24141](#), [pr#22180](#), Mykola Golub)
- rbd: [rbd-mirror] potential races during PoolReplayer shut-down ([issue#24008](#), [pr#22172](#), Jason Dillaman)
- rgw: add buffering filter to compression for fetch\_remote\_obj ([issue#23547](#), [pr#21758](#), Casey Bodley)
- rgw: add configurable AWS-compat invalid range get behavior ([issue#24317](#), [pr#22302](#), Matt Benjamin)
- rgw: admin rest api shouldn't return error when getting user's stats if ([issue#23821](#), [pr#21661](#), Zhang Shaowen)
- rgw: Allow swift acls to be deleted ([issue#22897](#), [pr#22465](#), Marcus Watts)
- rgw: aws4 auth supports PutBucketRequestPayment ([issue#23803](#), [pr#21660](#), Casey Bodley)
- rgw: beast frontend can listen on multiple endpoints ([issue#22779](#), [pr#21568](#), Casey Bodley)
- rgw: Bucket lifecycles stick around after buckets are deleted ([issue#19632](#), [pr#22551](#), Wei Qiaomiao)
- rgw: Do not modify email if argument is not set ([issue#24142](#), [pr#22352](#), Volker Theile)
- rgw: do not reflect period if not current ([issue#22844](#), [pr#21735](#), Tianshan Qu)
- rgw: es module: set compression type correctly ([issue#22758](#), [pr#21736](#), Abhishek Lekshmanan)
- rgw\_file: conditionally unlink handles when direct deleted ([issue#23299](#), [pr#21438](#), Matt Benjamin)
- rgw: fix bi\_list to reset is\_truncated flag if it skips entires ([issue#22721](#), [pr#21669](#), Orit Wasserman)

- rgw: fix 'copy part' without 'x-amz-copy-source-range' when compress... ([issue#23196](#), [pr#22438](#), fang yuxiang)
- rgw: fix error handling for GET with ?torrent ([issue#23506](#), [pr#21674](#), Casey Bodley)
- rgw: fix use of libcurl with empty header values ([issue#23663](#), [pr#21738](#), Casey Bodley)
- rgw:lc: RGWPutLC return ERR\_MALFORMED\_XML when missing <Rule> tag in... ([issue#21377](#), [pr#19884](#), Shasha Lu)
- rgw: making implicit\_tenants backwards compatible ([issue#24348](#), [pr#22363](#), Marcus Watts)
- rgw: Misnamed S3 operation ([issue#24061](#), [pr#21917](#), xiangxiang)
- rgw: move all pool creation into rgw\_init\_ioctx ([issue#23480](#), [pr#21675](#), Casey Bodley)
- rgw: radosgw-admin should not use metadata cache for readonly commands ([issue#23468](#), [pr#21437](#), Orit Wasserman)
- rgw: raise log level on coroutine shutdown errors ([issue#23974](#), [pr#21792](#), Casey Bodley)
- rgw: return EINVAL if max\_keys can not convert correctly ([issue#23586](#), [pr#21435](#), yuliyang)
- rgw: rgw\_statfs should report the correct stats ([issue#22202](#), [pr#21724](#), Supriti Singh)
- rgw: trim all spaces inside a metadata value ([issue#23301](#), [pr#22177](#), Orit Wasserman)
- slow mon ops from osd\_failure ([issue#24322](#), [pr#22568](#), Sage Weil)
- table of contents doesn't render for luminous/jewel docs ([issue#23780](#), [pr#21502](#), Alfredo Deza)
- test/librados: increase pgp\_num along with pg\_num ([issue#23763](#), [pr#21556](#), Kefu Chai)
- test/rgw: fix for bucket checkpoints ([issue#24212](#), [pr#22541](#), Casey Bodley)
- tests: filestore journal replay does not guard omap operations ([issue#22920](#), [pr#21547](#), Sage Weil)
- tools: ceph-disk: write log to /var/log/ceph not to /var/run/ceph ([issue#24041](#), [pr#21870](#), Kefu Chai)
- tools: ceph-fuse: getgroups failure causes exception ([issue#23446](#), [pr#21687](#), Jeff

Layton)

## v12.2.5 Luminous

This is the fifth bugfix release of Luminous v12.2.x long term stable release series. This release contains a range of bug fixes across all components of Ceph. We recommend all the users of 12.2.x series to update.

## Notable Changes

- MGR

The ceph-rest-api command-line tool included in the ceph-mon package has been obsoleted by the MGR “restful” module. The ceph-rest-api tool is hereby declared deprecated and will be dropped in Mimic.

The MGR “restful” module provides similar functionality via a “pass through” method. See <http://docs.ceph.com/docs/luminous/mgr/restful> for details.

- CephFS

Upgrading an MDS cluster to 12.2.3+ will result in all active MDS exiting due to feature incompatibilities once an upgraded MDS comes online (even as standby). Operators may ignore the error messages and continue upgrading/restarting or follow this upgrade sequence:

Reduce the number of ranks to 1 (`ceph fs set <fs_name> max_mds 1`), wait for all other MDS to deactivate, leaving the one active MDS, upgrade the single active MDS, then upgrade/start standbys. Finally, restore the previous `max_mds`.

See also: <https://tracker.ceph.com/issues/23172>

## Other Notable Changes

- add `-add-bucket` and `-move` options to `crushtool` ([issue#23472](#), [issue#23471](#), [pr#21079](#), Kefu Chai)
- `BlueStore.cc: _balance_bluefs_freespace: assert(0 == "allocate failed, wtf")` ([issue#23063](#), [pr#21394](#), Igor Fedotov, xie xingguo, Sage Weil, Zac Medico)
- `bluestore: correctly check all block devices to decide if journal is...` ([issue#23173](#), [issue#23141](#), [pr#20651](#), Greg Farnum)
- `bluestore: statfs available can go negative` ([issue#23074](#), [pr#20554](#), Igor Fedotov, Sage Weil)
- build Debian installation packages failure ([issue#22856](#), [issue#22828](#), [pr#20250](#),

Tone Zhang)

- build/ops: deb: move python-jinja2 dependency to mgr ([issue#22457](#), [pr#20748](#), Nathan Cutler)
- build/ops: deb: move python-jinja2 dependency to mgr ([issue#22457](#), [pr#21233](#), Nathan Cutler)
- build/ops: run-make-check.sh: fix SUSE support ([issue#22875](#), [issue#23178](#), [pr#20737](#), Nathan Cutler)
- cephfs-journal-tool: Fix Dumper destroyed before shutdown ([issue#22862](#), [issue#22734](#), [pr#20251](#), dongdong tao)
- ceph.in: print all matched commands if arg missing ([issue#22344](#), [issue#23186](#), [pr#20664](#), Luo Kexue, Kefu Chai)
- ceph-objectstore-tool command to trim the pg log ([issue#23242](#), [pr#20803](#), Josh Durgin, David Zafman)
- ceph osd force-create-pg cause all ceph-mon to crash and unable to come up again ([issue#22942](#), [pr#20399](#), Sage Weil)
- ceph-volume: adds raw device support to 'lvm list' ([issue#23140](#), [pr#20647](#), Andrew Schoen)
- ceph-volume: allow parallel creates ([issue#23757](#), [pr#21509](#), Theofilos Mouratidis)
- ceph-volume: allow skipping systemd interactions on activate/create ([issue#23678](#), [pr#21538](#), Alfredo Deza)
- ceph-volume: automatic VDO detection ([issue#23581](#), [pr#21505](#), Alfredo Deza)
- ceph-volume be resilient to \$PATH issues ([pr#20716](#), Alfredo Deza)
- ceph-volume: fix action plugins path in tox ([pr#20923](#), Guillaume Abrioux)
- ceph-volume Implement an 'activate all' to help with dense servers or migrating OSDs ([pr#21533](#), Alfredo Deza)
- ceph-volume improve robustness when reloading vms in tests ([pr#21072](#), Alfredo Deza)
- ceph-volume lvm.activate error if no bluestore OSDs are found ([issue#23644](#), [pr#21335](#), Alfredo Deza)
- ceph-volume: Nits noticed while studying code ([pr#21565](#), Dan Mick)
- ceph-volume tests alleviate libvirt timeouts when reloading ([issue#23163](#), [pr#20754](#), Alfredo Deza)
- ceph-volume update man page for prepare/activate flags ([pr#21574](#), Alfredo Deza)

- ceph-volume: Using -readonly for {vg|pv|lv}s commands ([pr#21519](#), Erwan Velu)
- client: allow client to use caps that are revoked but not yet returned ([issue#23028](#), [issue#23314](#), [pr#20904](#), Jeff Layton)
- : Client:Fix readdir bug ([issue#22936](#), [pr#20356](#), dongdong tao)
- client: release revoking Fc after invalidate cache ([issue#22652](#), [pr#20342](#), "Yan, Zheng")
- Client: setattr should drop "Fs" rather than "As" for mtime and size ([issue#22935](#), [pr#20354](#), dongdong tao)
- client: use either dentry\_invalidate\_cb or remount\_cb to invalidate k... ([issue#23355](#), [pr#20960](#), Zhi Zhang)
- cls/rbd: group\_image\_list incorrectly flagged as RW ([issue#23407](#), [issue#23388](#), [pr#20967](#), Jason Dillaman)
- cls/rgw: fix bi\_log\_iterate\_entries return wrong truncated ([issue#22737](#), [issue#23225](#), [pr#21054](#), Tianshan Qu)
- cmake: rbd resource agent needs to be executable ([issue#22980](#), [pr#20617](#), Tim Bishop)
- common/dns\_resolv.cc: Query for AAAA-record if ms\_bind\_ipv6 is True ([issue#23078](#), [issue#23174](#), [pr#20710](#), Wido den Hollander)
- common/ipaddr: Do not select link-local IPv6 addresses ([issue#21813](#), [pr#21111](#), Willem Jan Withagen)
- common: omit short option for id in help for clients ([issue#23156](#), [issue#23041](#), [pr#20654](#), Patrick Donnelly)
- common: should not check for VERSION\_ID ([issue#23477](#), [issue#23478](#), [pr#21090](#), Kefu Chai, Shengjing Zhu)
- config: Change bluestore\_cache\_kv\_max to type INT64 ([pr#20334](#), Zhi Zhang)
- Couldn't init storage provider (RADOS) ([issue#23349](#), [issue#22351](#), [pr#20896](#), Brad Hubbard)
- doc: Add missing pg states from doc ([issue#23113](#), [pr#20584](#), David Zafman)
- doc: outline upgrade procedure for mds cluster ([issue#23634](#), [issue#23568](#), [pr#21352](#), Patrick Donnelly)
- doc/rgw: add page for http frontend configuration ([issue#13523](#), [issue#22884](#), [pr#20242](#), Casey Bodley)
- doc: rgw: mention the civetweb support for binding to multiple ports ([issue#20942](#), [issue#23317](#), [pr#20906](#), Abhishek Lekshmanan)

- docs fix ceph-volume missing sub-commands ([pr#20691](#), Katie Holly, Yao Zongyou, David Galloway, Sage Weil, Alfredo Deza)
- doc: update man page to explain ceph-volume support bluestore ([issue#23142](#), [issue#22663](#), [pr#20679](#), lijing)
- Double free in rados\_getxattrs\_next ([issue#22940](#), [issue#22042](#), [pr#20358](#), Gu Zhongyan)
- fixes for openssl & libcurl ([issue#23239](#), [issue#23245](#), [issue#22951](#), [issue#23221](#), [issue#23203](#), [pr#20722](#), Marcus Watts, Abhishek Lekshmanan, Jesse Williamson)
- invalid JSON returned when querying pool parameters ([issue#23312](#), [issue#23200](#), [pr#20890](#), Chang Liu)
- is\_qemu\_running in qemu\_rebuild\_object\_map.sh and qemu\_dynamic\_features.sh may return false positive ([issue#23524](#), [pr#21192](#), Mykola Golub)
- [journal] allocating a new tag after acquiring the lock should use on-disk committed position ([issue#23011](#), [issue#22945](#), [pr#20454](#), Jason Dillaman)
- journal: Message too long error when appending journal ([issue#23545](#), [issue#23526](#), [pr#21216](#), Mykola Golub)
- legal: remove doc license ambiguity ([issue#23410](#), [issue#23336](#), [pr#20988](#), Nathan Cutler)
- librados: make OPERATION\_FULL\_FORCE the default for rados\_remove() ([issue#23114](#), [issue#22413](#), [pr#20585](#), Kefu Chai)
- librados/snap\_set\_diff: don't assert on empty snapset ([issue#23423](#), [pr#20991](#), Mykola Golub)
- librbd: potential crash if object map check encounters error ([issue#22857](#), [issue#22819](#), [pr#20253](#), Jason Dillaman)
- log: Fix AddressSanitizer: new-delete-type-mismatch ([issue#23324](#), [issue#23412](#), [pr#20998](#), Brad Hubbard)
- mds: add uptime to MDS status ([issue#23150](#), [pr#20626](#), Patrick Donnelly)
- mds: FAILED assert (p != active\_requests.end()) in MDRequestRef MDCache::request\_get(metareqid\_t) ([issue#23154](#), [issue#23059](#), [pr#21176](#), "Yan, Zheng")
- mds: fix session reference leak ([issue#22821](#), [issue#22969](#), [pr#20432](#), "Yan, Zheng")
- mds: optimize getattr file size ([issue#23013](#), [issue#22925](#), [pr#20455](#), "Yan, Zheng")

- mgr: Backport recent prometheus exporter changes ([pr#20642](#), Jan Fajerski, Boris Ranto)
- mgr: Backport recent prometheus rgw changes ([pr#21492](#), Jan Fajerski, John Spray, Boris Ranto, Rubab-Syed)
- mgr/balancer: pool-specific optimization support and bug fixes ([pr#20359](#), xie xingguo)
- mgr: die on bind() failure ([issue#23175](#), [pr#20712](#), John Spray)
- mgr: fix MSG\_MGR\_MAP handling ([issue#23409](#), [pr#20973](#), Gu Zhongyan)
- mgr: prometheus: fix PG state names ([pr#21365](#), John Spray)
- mgr: prometheus: set metadata metrics value to '1' (#22717) ([pr#20254](#), Konstantin Shalygin)
- mgr: quieten logging on missing OSD stats ([issue#23224](#), [pr#21053](#), John Spray)
- mgr/zabbix: Backports to Luminous ([pr#20781](#), Wido den Hollander)
- mon: allow removal of tier of ec overwritable pool ([issue#22971](#), [issue#22754](#), [pr#20433](#), Patrick Donnelly)
- mon: ops get stuck in "resend forwarded message to leader" ([issue#22114](#), [issue#23077](#), [pr#21016](#), Kefu Chai, Greg Farnum)
- mon, osd: fix potential collided \*Up Set\* after PG remapping ([issue#23118](#), [pr#20829](#), xie xingguo)
- mon/OSDMonitor.cc: fix expected\_num\_objects interpret error ([issue#22530](#), [issue#23315](#), [pr#20907](#), Yang Honggang)
- mon: update PaxosService::cached\_first\_committed in PaxosService::may... ([issue#23626](#), [issue#11332](#), [pr#21328](#), Xuehan Xu, yupeng chen)
- msg/async: size of EventCenter::file\_events should be greater than fd ([issue#23253](#), [issue#23306](#), [pr#20867](#), Yupeng Chen)
- Objecter: add ignore overlay flag if got redirect reply ([pr#20766](#), Ting Yi Lin)
- os/bluestore: avoid overhead of std::function in blob\_t ([pr#20674](#), Radoslaw Zarzynski)
- os/bluestore: avoid unneeded BlobRefing in \_do\_read() ([pr#20675](#), Radoslaw Zarzynski)
- os/bluestore: backport fixes around \_reap\_collection ([pr#20964](#), Jianpeng Ma)
- os/bluestore: change the type of aio\_t::res to long ([issue#23527](#), [issue#23544](#), [pr#21231](#), kungf)

- os/bluestore: \_dump\_onode() don't prolongate Onode anymore ([pr#20676](#), Radoslaw Zarzynski)
- os/bluestore: recalc\_allocated() when decoding bluefs\_fnode\_t ([issue#23256](#), [issue#23212](#), [pr#20771](#), Jianpeng Ma, Igor Fedotov, Kefu Chai)
- os/bluestore: trim cache every 50ms (instead of 200ms) ([issue#23226](#), [pr#21059](#), Sage Weil)
- osd: add numpg\_removing metric ([pr#20785](#), Sage Weil)
- osdc/Journaler: make sure flush() writes enough data ([issue#22967](#), [issue#22824](#), [pr#20431](#), "Yan, Zheng")
- osd: do not release\_reserved\_pushes when requeueing ([pr#21229](#), Sage Weil)
- osd: Fix assert when checking missing version ([issue#21218](#), [issue#23024](#), [pr#20495](#), David Zafman)
- osd: objecter sends out of sync with pg epochs for proxied ops ([issue#22123](#), [issue#23075](#), [pr#20609](#), Sage Weil)
- osd/OSDMap: skip out/crush-out osds ([pr#20840](#), xie xingguo)
- osd/osd\_types: fix pg\_pool\_t encoding for hammer ([pr#21283](#), Sage Weil)
- osd: remove cost from mclock op queues; cost not handled well in dmcl... ([pr#21426](#), J. Eric Ivancich)
- osd: Remove partially created pg known as DNE ([issue#23160](#), [issue#21833](#), [pr#20668](#), David Zafman)
- osd: resend osd\_pgtemp if it's not acked ([issue#23610](#), [issue#23630](#), [pr#21330](#), Kefu Chai)
- osd: treat successful and erroneous writes the same for log trimming ([issue#23323](#), [issue#22050](#), [pr#20851](#), Josh Durgin)
- os/filestore: fix do\_copy\_range replay bug ([issue#23351](#), [issue#23298](#), [pr#20957](#), Sage Weil)
- parent blocks are still seen after a whole-object discard ([issue#23304](#), [issue#23285](#), [pr#20860](#), Ilya Dryomov, Jason Dillaman)
- PendingReleaseNotes: add note about upgrading MDS ([issue#23414](#), [pr#21001](#), Patrick Donnelly)
- : qa: adjust cephfs full test for kclient ([issue#22966](#), [issue#22886](#), [pr#20417](#), "Yan, Zheng")
- qa: ignore io pause warnings in mds-full test ([issue#23062](#), [issue#22990](#), [pr#20525](#), Patrick Donnelly)

- qa: ignore MON\_DOWN while thrashing mons ([issue#23061](#), [pr#20523](#), Patrick Donnelly)
- qa/rgw: remove some civetweb overrides for beast testing ([issue#23002](#), [issue#23176](#), [pr#20736](#), Casey Bodley)
- qa: src/test/libcephfs/test.cc:376: Expected: (len) > (0), actual: -34 vs 0 ([issue#22383](#), [issue#22221](#), [pr#21173](#), Patrick Donnelly)
- qa: synchronize kcephfs suites with fs/multimds ([issue#22891](#), [issue#22627](#), [pr#20302](#), Patrick Donnelly)
- qa/tests - added tag: v12.2.2 to be used by client.1 ([pr#21452](#), Yuri Weinstein)
- qa/tests - Change machine type from 'vps' to 'ovh' as 'vps' does not ... ([pr#21031](#), Yuri Weinstein)
- qa/workunits/rados/test-upgrade-to-mimic.sh: fix tee output ([pr#21506](#), Sage Weil)
- qa/workunits/rbd: switch devstack tempest to 17.2.0 tag ([issue#23177](#), [issue#22961](#), [pr#20715](#), Jason Dillaman)
- radosgw-admin data sync run crashes ([issue#23180](#), [pr#20762](#), lvshanchun)
- rbd-mirror: fix potential infinite loop when formatting status message ([issue#22964](#), [issue#22932](#), [pr#20416](#), Mykola Golub)
- rbd-nbd: fix ebusy when do map ([issue#23542](#), [issue#23528](#), [pr#21230](#), Li Wang)
- rgw: add radosgw-admin sync error trim to trim sync error log ([issue#23302](#), [pr#20859](#), fang yuxiang)
- rgw: add xml output header in RGWCopyObj\_ObjStore\_S3 response msg ([issue#22416](#), [issue#22635](#), [pr#19883](#), Enming Zhang)
- rgw: Admin API Support for bucket quota change ([issue#23357](#), [issue#21811](#), [pr#20885](#), Jeegn Chen)
- rgw: allow beast frontend to listen on specific IP address ([issue#22858](#), [issue#22778](#), [pr#20252](#), Yuan Zhou)
- rgw: can't download object with range when compression enabled ([issue#23146](#), [issue#23179](#), [issue#22852](#), [pr#20741](#), fang yuxiang)
- rgw: data sync of versioned objects, note updating bi marker ([issue#23025](#), [pr#21214](#), Yehuda Sadeh)
- RGW doesn't check time skew in auth v4 http header request ([issue#23252](#), [issue#22766](#), [issue#22439](#), [issue#22418](#), [pr#20072](#), Bingyin Zhang, Casey Bodley)
- rgw\_file: avoid evaluating nullptr for readdir offset ([issue#22889](#), [pr#20345](#), Matt Benjamin)

- rgw: fix crash with rgw\_run\_sync\_thread false ([issue#23318](#), [issue#20448](#), [pr#20932](#), Orit Wasserman)
- rgw: fix memory fragmentation problem reading data from client ([issue#23347](#), [pr#20953](#), Marcus Watts)
- rgw: fix multisite read-write issues ([issue#23690](#), [issue#22804](#), [pr#21390](#), Niu Pengju)
- rgw: fix the max-uploads parameter not work ([issue#23020](#), [issue#22825](#), [pr#20476](#), Xin Liao)
- rgw\_log, rgw\_file: account for new required envvars ([issue#23192](#), [issue#21942](#), [pr#20672](#), Matt Benjamin)
- rgw: log the right http status code in civetweb frontend's access log ([issue#22812](#), [issue#22538](#), [pr#20157](#), Yao Zongyou)
- rgw: parse old rgw\_obj with namespace correctly ([issue#23102](#), [issue#22982](#), [pr#20586](#), Yehuda Sadeh)
- rgw recalculate stats option added ([issue#23691](#), [issue#23720](#), [issue#23335](#), [issue#23322](#), [pr#21393](#), Abhishek Lekshmanan)
- rgw: reject encrypted object COPY before supported ([issue#23232](#), [issue#23346](#), [pr#20937](#), Jeegn Chen)
- rgw: rgw: reshards cancel command should clear bucket resharding flag ([issue#21619](#), [pr#21389](#), Orit Wasserman)
- rgw: s3website error handler uses original object name ([issue#23201](#), [issue#23310](#), [pr#20889](#), Casey Bodley)
- rgw: update the max-buckets when the quota is uploaded ([issue#23022](#), [pr#20477](#), zhaokun)
- rgw: usage log fixes ([issue#23686](#), [issue#23758](#), [pr#21388](#), Yehuda Sadeh, Greg Farnum, Robin H. Johnson)
- rocksdb: incorporate the fix in RocksDB for no fast CRC32 path ([issue#22534](#), [pr#20825](#), Radoslaw Zarzynski)
- scrub errors not cleared on replicas can cause inconsistent pg state when replica takes over primary ([issue#23267](#), [pr#21103](#), David Zafman)
- snapmapper inconsistency, crash on luminous ([issue#23500](#), [pr#21118](#), Sage Weil)
- Special scrub handling of hinfo\_key errors ([issue#23654](#), [issue#23428](#), [issue#23364](#), [pr#21397](#), David Zafman)
- src: s/-use-wheel// ([pr#21177](#), Kefu Chai)

- systemd: Wait 10 seconds before restarting ceph-mgr ([issue#23083](#), [issue#23101](#), [pr#20604](#), Wido den Hollander)
- test\_admin\_socket.sh may fail on wait\_for\_clean ([issue#23507](#), [pr#21124](#), Mykola Golub)
- test/ceph-disk: specify the python used for creating venv ([issue#23281](#), [pr#20817](#), Kefu Chai)
- TestLibRBD.RenameViaLockOwner may still fail with -ENOENT ([issue#23152](#), [issue#23068](#), [pr#20628](#), Mykola Golub)
- test/librbd: utilize unique pool for cache tier testing ([issue#23064](#), [issue#11502](#), [pr#20550](#), Jason Dillaman)
- test/pybind/test\_rbd: allow v1 images for testing ([pr#21471](#), Sage Weil)
- test: Replace bc command with printf command ([pr#21015](#), David Zafman)
- tests: drop upgrade/jewel-x/point-to-point-x in luminous and master ([issue#23159](#), [issue#22888](#), [pr#20641](#), Nathan Cutler)
- tests: ENGINE Error in 'start' listener <bound in rados ([issue#23606](#), [pr#21307](#), John Spray)
- tests: rgw: swift tests target ceph-luminous branch ([pr#21048](#), Nathan Cutler)
- tests: unittest\_pglog timeout ([issue#23522](#), [issue#23504](#), [pr#21134](#), Nathan Cutler)
- Update mgr/restful documentation ([issue#23230](#), [pr#20725](#), Boris Ranto)

## v12.2.4 Luminous

---

This is the fourth bugfix release of Luminous v12.2.x long term stable release series. This was primarily intended to fix a few build, ceph-volume/ceph-disk and RGW issues. We recommend all the users of 12.2.x series to update.

## Notable Changes

---

- ceph-volume: adds support to zap encrypted devices ([issue#22878](#), [pr#20545](#), Andrew Schoen)
- ceph-volume: log the current running command for easier debugging ([issue#23004](#), [pr#20597](#), Andrew Schoen)
- ceph-volume: warn on mix of filestore and bluestore flags ([issue#23003](#), [pr#20568](#), Alfredo Deza)
- cmake: check bootstrap.sh instead before downloading boost ([issue#23071](#),

[pr#20515](#), Kefu Chai)

- core: Backport of cache manipulation: issues #22603 and #22604 ([issue#22604](#), [issue#22603](#), [pr#20353](#), Adam C. Emerson)
- core: last-stat-seq returns 0 because osd stats are cleared ([issue#23093](#), [pr#20548](#), Sage Weil, David Zafman)
- core: Snapset inconsistency is detected with its own error ([issue#22996](#), [pr#20501](#), David Zafman)
- rgw: make init env methods return an error ([issue#23039](#), [pr#20564](#), Abhishek Lekshmanan)
- rgw: parse old rgw\_obj with namespace correctly ([issue#22982](#), [pr#20566](#), Yehuda Sadeh)
- rgw: return valid Location element, CompleteMultipartUpload ([issue#22655](#), [pr#20266](#), Matt Benjamin)
- rgw: URL-decode S3 and Swift object-copy URLs ([issue#22121](#), [issue#22729](#), [pr#20236](#), Malcolm Lee, Matt Benjamin)
- rgw: use explicit index pool placement ([issue#22928](#), [pr#20565](#), Yehuda Sadeh)
- tools: ceph-disk: v12.2.2 unable to create bluestore osd using ceph-disk ([issue#22354](#), [pr#20563](#), Kefu Chai)
- tools: ceph-objectstore-tool: “\$OBJ get-omaphdr” and “\$OBJ list-omap” scan all pgs instead of using specific pg ([issue#21327](#), [pr#20283](#), David Zafman)

## v12.2.3 Luminous

---

This is the third bugfix release of Luminous v12.2.x long term stable release series. It contains a range of bug fixes and a few features across Bluestore, CephFS, RBD & RGW. We recommend all the users of 12.2.x series update.

## Notable Changes

---

- *CephFS*:
  - The CephFS client now checks for older kernels’ inability to reliably clear dentries from the kernel dentry cache. The new option `client_die_on_failed_dentry_invalidate` (default: true) may be turned off to allow the client to proceed (dangerous!).

## Other Notable Changes

---

- bluestore: do not crash on over-large objects ([issue#22161](#), [pr#19630](#), Sage Weil)
- bluestore: OSD crash on boot with assert caused by Bluefs on flush write ([issue#21932](#), [pr#19047](#), Jianpeng Ma)
- build/ops: ceph-base symbols not stripped in debs ([issue#22640](#), [pr#19969](#), Sage Weil)
- build/ops: ceph-conf: dump parsed config in plain text or as json ([issue#21862](#), [pr#18842](#), Piotr Dałek)
- build/ops: ceph-mgr dashboard has dependency on python-jinja2 ([issue#22457](#), [pr#19865](#), John Spray)
- build/ops: ceph-volume fails when centos7 image doesn't have lvm2 installed ([issue#22443](#), [issue#22217](#), [pr#20215](#), Nathan Cutler, Theofilos Mouratidis)
- build/ops: Default kernel.pid\_max is easily exceeded during recovery on high OSD-count system ([issue#21929](#), [pr#19133](#), David Disseldorp, Kefu Chai)
- build/ops: install-deps.sh: revert gcc to the one shipped by distro ([issue#22220](#), [pr#19680](#), Kefu Chai)
- build/ops: luminous build fails with -without-radosgw ([issue#22321](#), [pr#19483](#), Jason Dillaman)
- build/ops: move ceph-\*-tool binaries out of ceph-test subpackage ([issue#22319](#), [issue#21762](#), [pr#19355](#), liuchang0812, Nathan Cutler, Kefu Chai, Sage Weil)
- build.ops: rpm: adjust ceph-{osdomap,kvstore,monstore}-tool feature move ([issue#22558](#), [pr#19839](#), Kefu Chai)
- ceph: cluster [ERR] Unhandled exception from module 'balancer' while running on mgr.x: 'NoneType' object has no attribute 'iteritems'" in cluster log ([issue#22090](#), [pr#19023](#), Sage Weil)
- cephfs: cephfs-journal-tool: add "set pool\_id" option ([issue#22631](#), [pr#20085](#), dongdong tao)
- cephfs: cephfs-journal-tool: tool would miss to report some invalid range ([issue#22459](#), [pr#19626](#), dongdong tao)
- cephfs: cephfs: potential adjust failure in lru\_expire ([issue#22458](#), [pr#19627](#), dongdong tao)
- cephfs: "ceph tell mds" commands result in "File exists" errors on client admin socket ([issue#21406](#), [issue#21967](#), [pr#18831](#), Patrick Donnelly)
- cephfs: client: anchor Inode while trimming caps ([issue#22157](#), [pr#19105](#), Patrick Donnelly)

- cephfs: client: avoid recursive lock in ll\_get\_vino ([issue#22629](#), [pr#20086](#), dongdong tao)
- cephfs: client: dual client segfault with racing ceph\_shutdown ([issue#21512](#), [issue#20988](#), [pr#20082](#), Jeff Layton)
- cephfs: client: implement delegation support in userland cephfs ([issue#18490](#), [pr#19480](#), Jeff Layton)
- cephfs: client: quit on failed remount during dentry invalidate test #19218 ([issue#22269](#), [pr#19370](#), Patrick Donnelly)
- cephfs: List of filesystems does not get refreshed after a filesystem deletion ([issue#21599](#), [pr#18730](#), John Spray)
- cephfs: MDS : Avoid the assert failure when the inode for the cap\_export from other... ([issue#22610](#), [pr#20300](#), Jianyu Li)
- cephfs: MDSMonitor: monitor gives constant "is now active in filesystem cephfs as rank" cluster log info messages ([issue#21959](#), [pr#19055](#), Patrick Donnelly)
- cephfs: racy is\_mounted() checks in libcephfs ([issue#21025](#), [pr#17875](#), Jeff Layton)
- cephfs: src/mds/MDCache.cc: 7421: FAILED assert(CInode::count() == inode\_map.size() + snap\_inode\_map.size()) ([issue#21928](#), [pr#18912](#), "Yan, Zheng")
- cephfs: vstart\_runner: fixes for recent cephfs changes ([issue#22526](#), [pr#19829](#), Patrick Donnelly)
- ceph-volume: adds a -destroy flag to ceph-volume lvm zap ([issue#22653](#), [pr#20240](#), Andrew Schoen)
- ceph-volume: adds success messages for lvm prepare/activate/create ([issue#22307](#), [pr#20238](#), Andrew Schoen)
- ceph-volume: dmcrypt support for lvm ([issue#22619](#), [pr#20241](#), Alfredo Deza)
- ceph-volume dmcrypt support for simple ([issue#22620](#), [pr#20350](#), Andrew Schoen, Alfredo Deza)
- ceph-volume: do not use -key during mkfs ([issue#22283](#), [pr#20244](#), Kefu Chai, Sage Weil)
- ceph-volume: fix usage of the -osd-id flag ([issue#22642](#), [issue#22836](#), [pr#20323](#), Andrew Schoen)
- ceph-volume Format correctly when vg/lv cannot be used ([issue#22299](#), [pr#19527](#), Alfredo Deza)
- ceph-volume handle inline comments in the ceph.conf file ([issue#22297](#), [pr#19532](#),

Alfredo Deza)

- ceph-volume: handle leading whitespace/tabs in ceph.conf ([issue#22280](#), [pr#19526](#), Alfredo Deza)
- ceph-volume: lvm zap will unmount osd paths used by zapped devices ([issue#22876](#), [pr#20438](#), Andrew Schoen)
- ceph-volume: removed the explicit use of sudo ([issue#22282](#), [pr#19525](#), Andrew Schoen)
- ceph-volume rollback on failed OSD prepare/create ([issue#22281](#), [pr#20237](#), Alfredo Deza)
- ceph-volume should be able to handle multiple LVM (VG/LV) tags ([issue#22305](#), [pr#19528](#), Alfredo Deza)
- ceph-volume use realpath when checking mounts ([issue#22988](#), [pr#20429](#), Alfredo Deza)
- ceph-volume: warn on missing ceph.conf file ([issue#22326](#), [pr#19530](#), Alfredo Deza)
- common: compute SimpleLRU's size with contents.size() instead of lru.... ([issue#22613](#), [pr#19977](#), Xuehan Xu)
- config: lower default omap entries recovered at once ([issue#21897](#), [pr#19928](#), Josh Durgin)
- core: backoff causes out of order op ([issue#21407](#), [pr#18747](#), Sage Weil)
- core: common/throttle: start using 64-bit values ([issue#22539](#), [pr#19995](#), Igor Fedotov)
- core: fix broken use of stream::rdbuf() ([issue#22715](#), [pr#20042](#), Sage Weil)
- core: possible deadlock in various maintenance operations ([issue#22120](#), [pr#19123](#), Jason Dillaman)
- core: \_read\_bdev\_label unable to decode label at offset ([issue#22285](#), [pr#20326](#), Sage Weil)
- core: rocksdb: fixes early metadata spill over to slow device in ([issue#22264](#), [pr#19257](#), Igor Fedotov)
- core: Various odd clog messages for mons ([issue#22082](#), [pr#19031](#), John Spray)
- crush: balancer crush-compat sends "foo" command ([issue#22361](#), [pr#19555](#), John Spray)
- doc: crush\_ruleset is invalid command in luminous ([issue#20559](#), [pr#19446](#), Nathan Cutler)

- doc: doc/rbd: tweaks for the LIO iSCSI gateway ([issue#21763](#), [pr#20213](#), Ashish Singh, Mike Christie, Jason Dillaman)
- doc: man page for mount.fuse.ceph ([issue#21539](#), [issue#22595](#), [pr#19449](#), Jos Collin)
- doc: misc fixes for CephFS best practices ([issue#22630](#), [pr#19858](#), Jos Collin)
- doc: remove region from “INSTALL CEPH OBJECT GATEWAY” ([issue#21610](#), [pr#18865](#), Orit Wasserman)
- doc: update Blacklisting and OSD epoch barrier ([issue#22542](#), [pr#19741](#), Jos Collin)
- librbd: cannot clone all image-metas if we have more than 64 key/value pairs ([issue#21814](#), [pr#19503](#), PCzhangPC)
- librbd: cannot copy all image-metas if we have more than 64 key/value pairs ([issue#21815](#), [pr#19504](#), PCzhangPC)
- librbd: compare and write against a clone can result in failure ([issue#20789](#), [pr#20211](#), Mykola Golub, Jason Dillaman)
- librbd: default to sparse-reads for any IO operation over 64K ([issue#21849](#), [pr#20208](#), Jason Dillaman)
- librbd: fix snap create/rm may taking long time ([issue#22716](#), [pr#20153](#), Song Shun)
- librbd: force removal of a snapshot cannot ignore dependent children ([issue#22791](#), [pr#20135](#), Jason Dillaman)
- librbd: Image-meta should be dynamically refreshed ([issue#21529](#), [pr#19447](#), Dongsheng Yang, Jason Dillaman)
- librbd: journal should ignore -EILSEQ errors from compare-and-write ([issue#21628](#), [pr#20206](#), Jason Dillaman)
- librbd: refresh image after applying new/removing old metadata ([issue#21711](#), [pr#19485](#), Jason Dillaman)
- librbd: set deleted parent pointer to null ([issue#22158](#), [pr#20210](#), Jason Dillaman)
- luminous: ceph-fuse: ::rmdir() uses a deleted memory structure of dentry leads ... ([issue#22536](#), [pr#19968](#), YunfeiGuan)
- mds: check for CEPH\_OSDMAP\_FULL is now wrong; cluster full flag is obsolete ([issue#22483](#), [pr#19830](#), Patrick Donnelly)
- mds: don't check gid when none specified in auth caps ([issue#22009](#), [pr#18835](#),

Douglas Fuller)

- mds: don't delay processing completed requests in replay queue ([issue#22163](#), [pr#19157](#), "Yan, Zheng")
- mds: don't report repaired backtraces in damagetable, write back after repair, clean up scrub log ([issue#18743](#), [issue#22058](#), [pr#20341](#), "Yan, Zheng", John Spray)
- mds: fix CDir::log\_mark\_dirty() ([issue#21584](#), [pr#18008](#), "Yan, Zheng")
- mds: fix dump last\_sent ([issue#22562](#), [pr#19959](#), dongdong tao)
- mds: fix MDS\_FEATURE\_INCOMPAT\_FILE\_LAYOUT\_V2 definition ([issue#21985](#), [pr#18782](#), "Yan, Zheng")
- mds: fix return value of MDCache::dump\_cache ([issue#22798](#), [pr#20121](#), "Yan, Zheng")
- mds: fix scrub crash ([issue#22730](#), [pr#20249](#), dongdong tao)
- mds: fix StrayManager::truncate() ([issue#21091](#), [pr#18019](#), "Yan, Zheng")
- mds: handle client reconnect gather race ([issue#22263](#), [pr#19326](#), "Yan, Zheng")
- mds: handle client session messages when mds is stopping ([issue#22460](#), [pr#19585](#), "Yan, Zheng")
- mds: handle 'inode gets queued for recovery multiple times' ([issue#22647](#), [pr#19982](#), "Yan, Zheng")
- mds: ignore export pin for unlinked directory ([issue#22219](#), [pr#19360](#), "Yan, Zheng")
- mds: limit size of subtree migration ([issue#21892](#), [pr#20339](#), "Yan, Zheng")
- mds: no assertion on inode being purging in find\_ino\_peers() ([issue#21722](#), [pr#18869](#), Zhi Zhang)
- mds: preserve order of requests during recovery of multimds cluster ([issue#21843](#), [pr#18871](#), "Yan, Zheng")
- mds: prevent filelock from being stuck at XSYN state ([issue#22008](#), [pr#20340](#), "Yan, Zheng")
- mds: properly eval locks after importing inode ([issue#22357](#), [pr#19646](#), "Yan, Zheng")
- mds: reduce debugging level for balancer messages ([issue#21853](#), [pr#19827](#), Patrick Donnelly)
- mds: respect mds\_client\_writeable\_range\_max\_inc\_objs config ([issue#22492](#), [pr#19776](#), "Yan, Zheng")

- mds: set higher priority for some perf counters ([issue#22776](#), [pr#20299](#), Shangzhong Zhu)
- mds: set PRIO\_USEFUL on num\_sessions counter ([issue#21927](#), [pr#18722](#), John Spray)
- mds: tell session ls returns vanilla EINVAL when MDS is not active ([issue#21991](#), [pr#19505](#), Jos Collin)
- mds: track dirty dentries in separate list ([issue#19578](#), [pr#19775](#), "Yan, Zheng")
- mds: trim 'N' log segments according to how many log segments are there ([issue#21975](#), [pr#18783](#), "Yan, Zheng")
- mgr: ceph-mgr spuriously reloading OSD metadata on map changes ([issue#21159](#), [pr#18732](#), Yanhu Cao)
- mgr: disconnect unregistered service daemon when report received ([issue#22286](#), [pr#20089](#), Jason Dillaman)
- mgr: KeyError: ('name',) in balancer rm ([issue#22470](#), [pr#19624](#), Dan van der Ster)
- mgr: Manager daemon x is unresponsive. No standby daemons available ([issue#21147](#), [pr#19501](#), Sage Weil)
- mgr: mgr/balancer/upmap\_max\_iterations must be cast to integer ([issue#22429](#), [pr#19553](#), Dan van der Ster)
- mgr: mgr/dashboard: added iSCSI IOPS/throughput metrics ([issue#21391](#), [pr#20209](#), Jason Dillaman)
- mgr: mgr/dashboard: Fix PG status coloring ([issue#22615](#), [pr#19844](#), Wido den Hollander)
- mgr: mgr/prometheus: add missing 'deep' state to PG\_STATES in ceph-mgr pro... ([issue#22116](#), [pr#19929](#), Jan Fajerski, Peter Woodman)
- mgr: mgr tests don't indicate failure if exception thrown from serve() ([issue#21999](#), [pr#18832](#), John Spray)
- mgr: mgr[zabbix] float division by zero (osd['kb'] = 0) ([issue#21904](#), [pr#19048](#), Ilja Slepnev)
- mgr: prometheus: added osd commit/apply latency metrics (#22718) ([issue#22718](#), [pr#20084](#), Konstantin Shalygin)
- mgr: pybind/mgr/dashboard: fix duplicated slash in html href ([issue#22851](#), [pr#20325](#), Shengjing Zhu)
- mgr: pybind/mgr/dashboard: fix reverse proxy support ([issue#22557](#), [pr#20182](#), Nick Erdmann, Kefu Chai)
- mgr: pybind/mgr/prometheus: fix metric type undef -> untyped ([issue#22313](#),

[pr#19834](#), Ilya Margolin)

- mgr: restarting active ceph-mgr cause glitches in bps and iops metrics ([issue#21773](#), [pr#18735](#), Aleksei Gutikov, Kefu Chai)
- mgr: Services reported with blank hostname ([issue#20887](#), [issue#21687](#), [pr#17869](#), liuchang0812, Chang Liu)
- mon: do not use per\_pool\_sum\_delta to show recovery summary ([issue#22727](#), [pr#20150](#), Chang Liu)
- mon: fix mgr using auth\_client\_required policy ([issue#22096](#), [pr#20156](#), John Spray)
- mon: MDSMonitor: reject misconfigured mds\_blacklist\_interval ([issue#21821](#), [pr#19871](#), John Spray)
- mon/MgrMonitor: limit mgrmap history ([issue#22257](#), [pr#19187](#), Sage Weil)
- mon: reenable timer to send digest when paxos is temporarily inactive ([issue#22142](#), [pr#19481](#), Jan Fajerski)
- msg: msg/async/AsyncConnection.cc: 1835: FAILED assert(state == STATE\_CLOSED) ([issue#21883](#), [pr#18746](#), Haomai Wang)
- msg: msg/async: unregister connection failed when racing happened ([issue#22437](#), [pr#19552](#), Haomai Wang)
- osdc: “FAILED assert(bh->last\_write\_tid > tid)” in powercycle-wip-yuri-master-1.19.18-distro-basic-smithi ([issue#22741](#), [pr#20256](#), “Yan, Zheng”)
- osdc/Journaler: add ‘stopping’ check to various finish callbacks ([issue#22360](#), [pr#19610](#), “Yan, Zheng”)
- osdc/Objecter: objecter op\_send\_bytes perf counter always 0 ([issue#21982](#), [pr#19046](#), Jianpeng Ma)
- osd: do not check out-of-date osdmap for DESTROYED flag on start ([issue#22673](#), [pr#20068](#), Sage Weil)
- osd,mgr: report pending creating pg to mgr ([issue#22440](#), [pr#20204](#), Kefu Chai)
- osd: miscounting degraded objects and PG stuck in recovery\_unfound ([issue#22145](#), [pr#20055](#), Sage Weil, David Zafman)
- osd: Objecter::C\_ObjectOperation\_sparse\_read throws/catches exceptions on -ENOENT ([issue#21844](#), [pr#18744](#), Jason Dillaman)
- osd: Objecter::\_send\_op unnecessarily constructs costly hobject\_t ([issue#21845](#), [pr#18745](#), Jason Dillaman)
- osd: On pg repair the primary is not favored as was intended ([issue#21907](#),

[pr#19083](#), David Zafman)

- osd: OSD crushes with FAILED assert(used\_blocks.size() > count) during the first start after upgrade 12.2.1 -> 12.2.2 ([issue#22535](#), [pr#19888](#), Igor Fedotov)
- osd: OSDMap cache assert on shutdown ([issue#21737](#), [pr#18749](#), Greg Farnum)
- osd: OSDService::recovery\_need\_sleep read+updated without locking ([issue#21566](#), [pr#18753](#), Neha Ojha)
- osd: "osd status" command exception if OSD not in pgmap stats ([issue#21707](#), [pr#19084](#), Yanhu Cao)
- osd, pg, mgr: make snap trim queue problems visible ([issue#22448](#), [pr#20098](#), Piotr Dałek)
- osd: Pool Compression type option doesn't apply to new OSDs ([issue#22419](#), [pr#20106](#), Kefu Chai)
- osd: replica read can trigger cache promotion ([issue#20919](#), [pr#19499](#), Sage Weil)
- osd/ReplicatedPG.cc: recover\_replicas: object added to missing set for backfill, but is not in recovering, error! ([issue#21382](#), [issue#14513](#), [issue#18162](#), [pr#20081](#), David Zafman)
- osd: subscribe osdmmaps if any pending pgs ([issue#22113](#), [pr#19059](#), Kefu Chai)
- osd: "sudo cp /var/lib/ceph/osd/ceph-0/fsid ..." fails ([issue#20736](#), [pr#19631](#), Patrick Donnelly)
- os: fix 0-length zero semantics, test ([issue#21712](#), [pr#20049](#), Sage Weil)
- qa/tests: Applied PR 20053 to stress-split tests ([issue#22665](#), [pr#20451](#), Yuri Weinstein)
- rbd: abort in listing mapped nbd devices when running in a container ([issue#22012](#), [issue#22011](#), [pr#19051](#), Li Wang)
- rbd: [api] compare-and-write methods not properly advertised ([issue#22036](#), [pr#18834](#), Jason Dillaman)
- rbd: class rbd.Image discard--OSError: [errno 2147483648] error discarding region ([issue#21966](#), [pr#19058](#), Jason Dillaman)
- rbd: cluster resource agent ocf:ceph:rbd - wrong permissions ([issue#22362](#), [pr#19554](#), Nathan Cutler)
- rbd: disk usage on empty pool no longer returns an error message ([issue#22200](#), [pr#19107](#), Jason Dillaman)
- rbd: fix crash during map ([issue#21808](#), [pr#18698](#), Peter Keresztes Schmidt)

- rbd: [journal] tags are not being expired if no other clients are registered ([issue#21960](#), [pr#18840](#), Jason Dillaman)
- rbd: librbd: filter out potential race with image rename ([issue#18435](#), [pr#19853](#), Jason Dillaman)
- rbd-mirror: Allow a different data-pool to be used on the secondary cluster ([issue#21088](#), [pr#19305](#), Adam Wolfe Gordon)
- rbd-mirror: primary image should register in remote, non-primary image's journal ([issue#21961](#), [issue#21561](#), [pr#20207](#), Jason Dillaman)
- rbd-mirror: sync image metadata when transferring remote image ([issue#21535](#), [pr#19484](#), Jason Dillaman)
- rbd: Python RBD metadata\_get does not work ([issue#22306](#), [pr#19479](#), Mykola Golub)
- rbd: rbd ls -l crashes with SIGABRT ([issue#21558](#), [pr#19800](#), Jason Dillaman)
- rbd: [rbd-mirror] new pools might not be detected ([issue#22461](#), [pr#19625](#), Jason Dillaman)
- rbd: [rbd-nbd] Fedora does not register resize events ([issue#22131](#), [pr#19066](#), Jason Dillaman)
- rbd: [test] UpdateFeatures RPC message should be included in test\_notify.py ([issue#21936](#), [pr#18838](#), Jason Dillaman)
- Revert " luminous: msg/asynchronous: unregister connection failed when racing happened" ([issue#22231](#), [pr#20247](#), Sage Weil)
- rgw: 501 is returned when init multipart is using V4 signature and chunk encoding ([issue#22129](#), [pr#19506](#), Jeegn Chen)
- rgw: add cors header rule check in cors option request ([issue#22002](#), [pr#19053](#), yuliyang)
- rgw: backport beast frontend and boost 1.66 update ([issue#22101](#), [issue#20935](#), [issue#21831](#), [issue#20048](#), [issue#22600](#), [issue#20971](#), [pr#19848](#), Casey Bodley, Jiaying Ren)
- rgw: bucket index object not deleted after radosgw-admin bucket rm --purge-objects --bypass-gc ([issue#22122](#), [issue#19959](#), [pr#19085](#), Aleksei Gutikov)
- rgw: bucket policy evaluation logical error ([issue#21901](#), [issue#21896](#), [pr#19810](#), Adam C. Emerson)
- rgw: bucket resharding should not update bucket ACL or user stats ([issue#22742](#), [issue#22124](#), [pr#20327](#), Orit Wasserman)
- rgw: check going\_down() when lifecycle processing ([issue#22099](#), [pr#19088](#), Yao)

Zongyou)

- rgw: Dynamic bucket indexing, resharding and tenants seems to be broken ([issue#22046](#), [pr#19050](#), Orit Wasserman)
- rgw: file deadlock on lru evicting ([issue#22736](#), [pr#20075](#), Matt Benjamin)
- rgw: fix chained cache invalidation to prevent cache size growth ([issue#22410](#), [pr#19785](#), Mark Kogan)
- rgw: fix for empty query string in beast frontend ([issue#22797](#), [pr#20338](#), Casey Bodley)
- rgw: fix GET website response error code ([issue#22272](#), [pr#19489](#), Dmitry Plyakin)
- rgw: fix rewrite a versioning object create a new object bug ([issue#21984](#), [issue#22529](#), [pr#19787](#), Enming Zhang, Matt Benjamin)
- rgw: Fix swift object expiry not deleting objects ([issue#22084](#), [pr#18972](#), Pavan Rallabhandi)
- rgw: Fix swift object expiry not deleting objects ([issue#22084](#), [pr#19090](#), Pavan Rallabhandi)
- rgw: librgw: fix shutdown error with resources uncleared ([issue#22296](#), [pr#20073](#), Tao Chen)
- rgw: log keystone errors at a higher level ([issue#22151](#), [pr#19077](#), Abhishek Lekshmanan)
- rgw: make HTTP dechunking compatible with Amazon S3 ([issue#21015](#), [pr#19500](#), Radoslaw Zarzynski)
- rgw: modify s3 type subuser access permission fail ([issue#21983](#), [pr#18766](#), yuliyang)
- rgw: multisite: destination zone does not compress synced objects ([issue#21895](#), [pr#18867](#), Casey Bodley)
- rgw: multisite: 'radosgw-admin sync error list' contains temporary EBUSY errors ([issue#22473](#), [pr#19799](#), Casey Bodley)
- rgw: null instance mtime incorrect when enable versioning ([issue#21743](#), [pr#18870](#), Shasha Lu)
- rgw: Policy parser may or may not dereference uninitialized boost::optional sometimes ([issue#21962](#), [pr#18868](#), Adam C. Emerson)
- rgw: Possible deadlock in 'list\_children' when refresh is required ([issue#21670](#), [pr#18564](#), Jason Dillaman)
- rgw: put bucket policy panics RGW process ([issue#22541](#), [pr#19847](#), Bingyin Zhang)

- rgw: radosgw-admin reshards command argument error ([issue#21723](#), [pr#19502](#), Yao Zongyou)
- rgw: radosgw-admin zonegroup get and zone get should return defaults when there is no realm ([issue#21615](#), [pr#19086](#), lvshanchun)
- rgw: Random 500 errors in Swift PutObject (needs cache fixes) ([issue#22517](#), [issue#21560](#), [pr#19788](#), Adam C. Emerson)
- rgw: refuses upload when Content-Type missing from POST policy ([issue#20201](#), [pr#19867](#), Matt Benjamin)
- rgw: revert PR #8765 ([issue#22364](#), [pr#19434](#), fang.yuxiang)
- rgw: RGWCrashError: RGW will crash if a putting lc config request does not include an ID tag in the request xml ([issue#21980](#), [issue#22006](#), [pr#18765](#), Enming Zhang)
- rgw: rgw multisite: automated trimming for bucket index logs ([issue#18229](#), [pr#20062](#), Casey Bodley)
- rgw: RGW: S3 POST policy should not require Content-Type ([issue#20201](#), [pr#19784](#), Matt Benjamin)
- rgw: rgw segfaults after running radosgw-admin data sync init ([issue#22083](#), [pr#19071](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: rgw usage trim only trims a few entries ([issue#22234](#), [pr#19636](#), Abhishek Lekshmanan)
- rgw: S3 API Policy Conditions IpAddress and NotIpAddress do not work ([issue#20931](#), [issue#20991](#), [pr#19819](#), John Gibson, yuliyang, Casey Bodley, Abhishek Lekshmanan, Jiaying Ren)
- rgw: Segmentation fault when starting radosgw after reverting .rgw.root ([issue#21996](#), [pr#18764](#), Orit Wasserman, Casey Bodley)
- rgw: set sync\_from\_all as true when no value is seen ([issue#22062](#), [pr#19038](#), Abhishek Lekshmanan)
- rgw: unlink deleted bucket from bucket's owner ([issue#22248](#), [pr#20357](#), Casey Bodley)
- rgw: user stats increased after bucket reshards ([issue#22124](#), [pr#19538](#), Orit Wasserman)
- rgw: When a system object is created exclusively, do not distribute the ([issue#22792](#), [pr#20107](#), J. Eric Ivancich, Robin H. Johnson)
- tests: ceph\_test\_cls\_log failures related to cls\_cxx\_subop\_version() ([issue#21964](#), [pr#18715](#), Casey Bodley)

- tests: ceph\_test\_objectstore fails ObjectStore/StoreTest.Synthetic/1 (filestore) buffer content mismatch ([issue#21712](#), [issue#21818](#), [pr#18742](#), Sage Weil)
- tests: configure zabbix properly before selftest ([issue#22514](#), [pr#19831](#), John Spray)
- tests: do not configure ec data pool with memstore ([issue#22436](#), [pr#19628](#), Patrick Donnelly)
- tests: force backfill test can conflict with pool removal ([issue#22614](#), [pr#19966](#), Sage Weil)
- tests: full flag not set on osdmap for tasks.cephfs.test\_full ([issue#22475](#), [pr#19962](#), Patrick Donnelly)
- tests: increase osd count for ec testing ([issue#22646](#), [pr#19976](#), Patrick Donnelly)
- tests - Initial checkin for luminous point-to-point upgrade ([issue#22048](#), [pr#18771](#), Yuri Weinstein)
- tests: qa/workunits/rbd: simplify split-brain test to avoid potential race ([issue#22485](#), [pr#20205](#), Jason Dillaman)
- tests: qa/workunits/rbd: switch devstack to pike release ([issue#22786](#), [pr#20136](#), Jason Dillaman)
- tests: rbd\_mirror\_helpers.sh request\_resync\_image function saves image id to wrong variable ([issue#21663](#), [pr#19802](#), Jason Dillaman)
- tools/ceph\_monstore\_tool: include mgrmap in initial paxos epoch ([issue#22266](#), [pr#20116](#), Kefu Chai)
- tools: ceph-monstore-tool -readable mode doesn't understand FSMap, MgrMap ([issue#21577](#), [pr#18754](#), John Spray)
- tools: ceph-objectstore-tool: Add option dump-import to examine an export ([issue#22086](#), [pr#19487](#), David Zafman)
- tools: ceph\_objectstore\_tool: no flush before collection\_empty() calls; ObjectStore/StoreTest.SimpleAttrTest/2 fails ([issue#22409](#), [pr#19967](#), Igor Fedotov)
- tools: ceph-objectstore-tool set-size should clear data-digest ([issue#22112](#), [pr#20069](#), David Zafman)
- tools/crushtool: skip device id if no name exists ([issue#22117](#), [pr#19039](#), Jan Fajerski)

## v12.2.2 Luminous

This is the second bugfix release of Luminous v12.2.x long term stable release series. It contains a range of bug fixes and a few features across Bluestore, CephFS, RBD & RGW. We recommend all the users of 12.2.x series update.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- Standby ceph-mgr daemons now redirect requests to the active messenger, easing configuration for tools & users accessing the web dashboard, restful API, or other ceph-mgr module services.
- The prometheus module has several significant updates and improvements.
- The new balancer module enables automatic optimization of CRUSH weights to balance data across the cluster.
- The ceph-volume tool has been updated to include support for BlueStore as well as FileStore. The only major missing ceph-volume feature is dm-crypt support.
- RGW's dynamic bucket index resharding is disabled in multisite environments, as it can cause inconsistencies in replication of bucket indexes to remote sites.

## Other Notable Changes

- build/ops: bump sphinx to 1.6 ([issue#21717](#), [pr#18167](#), Kefu Chai, Alfredo Deza)
- build/ops: macros expanding in spec file comment ([issue#22250](#), [pr#19173](#), Ken Dreyer)
- build/ops: python-numpy-devel build dependency for SUSE ([issue#21176](#), [pr#17692](#), Nathan Cutler)
- build/ops: selinux: Allow setattr on lnk sysfs files ([issue#21492](#), [pr#18650](#), Boris Ranto)
- build/ops: Ubuntu amd64 client can not discover the ubuntu arm64 ceph cluster ([issue#19705](#), [pr#18293](#), Kefu Chai)
- core: buffer: fix ABI breakage by removing list \_mempool member ([issue#21573](#), [pr#18491](#), Sage Weil)
- core: Daemons(OSD, Mon...) exit abnormally at injectargs command ([issue#21365](#), [pr#17864](#), Yan Jun)
- core: Disable messenger logging (debug ms = 0/0) for clients unless overridden ([issue#21860](#), [pr#18529](#), Jason Dillaman)
- core: Improve OSD startup time by only scanning for omap corruption once ([issue#21328](#), [pr#17889](#), Luo Kexue, David Zafman)
- core: upmap does not respect osd reweights ([issue#21538](#), [pr#18699](#), Theofilos Mouratidis)
- dashboard: barfs on nulls where it expects numbers ([issue#21570](#), [pr#18728](#), John Spray)
- dashboard: OSD list has servers and osds in arbitrary order ([issue#21572](#), [pr#18736](#), John Spray)
- dashboard: the dashboard uses absolute links for filesystems and clients ([issue#20568](#), [pr#18737](#), Nick Erdmann)
- filestore: set default readahead and compaction threads for rocksdb ([issue#21505](#), [pr#18234](#), Josh Durgin, Mark Nelson)
- librbd: object map batch update might cause OSD suicide timeout ([issue#21797](#), [pr#18416](#), Jason Dillaman)
- librbd: snapshots should be created/removed against data pool ([issue#21567](#), [pr#18336](#), Jason Dillaman)

- mds: make sure snap inode's last matches its parent dentry's last ([issue#21337](#), [pr#17994](#), "Yan, Zheng")
- mds: sanitize mdsmap of removed pools ([issue#21945](#), [issue#21568](#), [pr#18628](#), Patrick Donnelly)
- mgr: bulk backport of ceph-mgr improvements ([issue#21594](#), [issue#17460](#), [issue#21197](#), [issue#21158](#), [issue#21593](#), [pr#18675](#), Benjeman Meekhof, Sage Weil, Jan Fajerski, John Spray, Kefu Chai, My Do, Spandan Kumar Sahu)
- mgr: ceph-mgr gets process called "exe" after respawn ([issue#21404](#), [pr#18738](#), John Spray)
- mgr: fix crashable DaemonStateIndex::get calls ([issue#17737](#), [pr#18412](#), John Spray)
- mgr: key mismatch for mgr after upgrade from jewel to luminous(dev) ([issue#20950](#), [pr#18727](#), John Spray)
- mgr: mgr status module uses base 10 units ([issue#21189](#), [issue#21752](#), [pr#18257](#), John Spray, Yanhu Cao)
- mgr: mgr[zabbix] float division by zero ([issue#21518](#), [pr#18734](#), John Spray)
- mgr: Prometheus crash when update ([issue#21253](#), [pr#17867](#), John Spray)
- mgr: prometheus module generates invalid output when counter names contain non-alphanum characters ([issue#20899](#), [pr#17868](#), John Spray, Jeremy H Austin)
- mgr: Quietен scary RuntimeError from restful module on startup ([issue#21292](#), [pr#17866](#), John Spray)
- mgr: Spurious ceph-mgr failovers during mon elections ([issue#20629](#), [pr#18726](#), John Spray)
- mon: Client client.admin marked osd.2 out, after it was down for 1504627577 seconds ([issue#21249](#), [pr#17862](#), John Spray)
- mon: DNS SRV default service name not used anymore ([issue#21204](#), [pr#17863](#), Kefu Chai)
- mon/MgrMonitor: handle cmd desc to/from disk in the absence of active mgr ([issue#21300](#), [pr#18038](#), Joao Eduardo Luis)
- mon/mgr: sync "mgr\_command\_descs","osd\_metadata" and "mgr\_metadata" prefixes to new mons ([issue#21527](#), [pr#18620](#), huanwen ren)
- mon: osd feature checks with 0 up osds ([issue#21471](#), [issue#20751](#), [pr#18364](#), Brad Hubbard, Sage Weil)
- mon,osd: fix "pg ls {forced\_backfill, backfilling}" ([issue#21609](#), [pr#18236](#), Kefu

Chai)

- mon/OSDMonitor: add option to fix up ruleset-\* to crush-\* for ec profiles ([issue#22128](#), [pr#18945](#), Sage Weil)
- mon, osd: per pool space-full flag support ([issue#21409](#), [pr#17730](#), xie xingguo)
- mon/PGMap: Fix %USED calculation ([issue#22247](#), [pr#19230](#), Xiaoxi Chen)
- mon: update get\_store\_prefixes implementations ([issue#21534](#), [pr#18621](#), John Spray, huanwen ren)
- msgr: messages/MOSDMap: do compat reencode of crush map, too ([issue#21882](#), [pr#18456](#), Sage Weil)
- msgr: src/messages/MOSDMap: reencode OSDMap for older clients ([issue#21660](#), [pr#18140](#), Sage Weil)
- os/bluestore/BlueFS: fix race with log flush during async log compaction ([issue#21878](#), [pr#18503](#), Sage Weil)
- os/bluestore: fix another aio stall/deadlock ([issue#21470](#), [pr#18127](#), Sage Weil)
- os/bluestore: fix SharedBlob unregistration ([issue#22039](#), [pr#18983](#), Sage Weil)
- os/bluestore: handle compressed extents in blob unsharing checks ([issue#21766](#), [pr#18501](#), Sage Weil)
- os/bluestore: replace 21089 repair with something online (instead of fsck) ([issue#21089](#), [pr#17734](#), Sage Weil)
- os/bluestore: set bitmap freelist resolution to min\_alloc\_size ([issue#21408](#), [pr#18050](#), Sage Weil)
- os/blueStore::umount will crash when the BlueStore is opened by start\_kv\_only() ([issue#21624](#), [pr#18750](#), Chang Liu)
- osd: additional protection for out-of-bounds EC reads ([issue#21629](#), [pr#18413](#), Jason Dillaman)
- osd: allow recovery preemption ([issue#21613](#), [pr#18025](#), Sage Weil)
- osd: build\_past\_intervals\_parallel: Ignore new partially created PGs ([issue#21833](#), [pr#18673](#), David Zafman)
- osd: dump bluestore debug on shutdown if debug option is set ([issue#21259](#), [pr#18103](#), Sage Weil)
- osd: make stat\_bytes and stat\_bytes\_used counters PRI0\_USEFUL ([issue#21981](#), [pr#18723](#), Yao Zongyou)
- osd: make the PG's SORTBITWISE assert a more generous shutdown ([issue#20416](#),

pr#18132, Greg Farnum)

- osd: OSD metadata ‘backend\_filestore\_dev\_node’ is unknown even for simple deployment ([issue#20944](#), [pr#17865](#), Sage Weil)
- rbd: [cli] mirror getter commands will fail if mirroring has never been enabled ([issue#21319](#), [pr#17861](#), Jason Dillaman)
- rbd: cls/journal: fixed possible infinite loop in expire\_tags ([issue#21956](#), [pr#18626](#), Jason Dillaman)
- rbd: cls/journal: possible infinite loop within tag\_list class method ([issue#21771](#), [pr#18417](#), Jason Dillaman)
- rbd: [rbd-mirror] asok hook names not updated when image is renamed ([issue#20860](#), [pr#17860](#), Mykola Golub)
- rbd: [rbd-mirror] forced promotion can result in incorrect status ([issue#21559](#), [pr#18337](#), Jason Dillaman)
- rbd: [rbd-mirror] peer cluster connections should filter out command line optionals ([issue#21894](#), [pr#18566](#), Jason Dillaman)
- rgw: add support for Swift’s per storage policy statistics ([issue#17932](#), [issue#21506](#), [pr#17835](#), Radoslaw Zarzynski, Casey Bodley)
- rgw: add support for Swift’s reversed account listings ([issue#21148](#), [pr#17834](#), Radoslaw Zarzynski)
- rgw: avoid logging keystone revocation failures when no keystone is configured ([issue#21400](#), [pr#18441](#), Abhishek Lekshmanan)
- rgw: disable dynamic resharding in multisite environment ([issue#21725](#), [pr#18432](#), Orit Wasserman)
- rgw: encryption: PutObj response does not include sse-kms headers ([issue#21576](#), [pr#18442](#), Casey Bodley)
- rgw: encryption: reject requests that don’t provide all expected headers ([issue#21581](#), [pr#18429](#), Enming Zhang)
- rgw: expose -sync-stats via admin api ([issue#21301](#), [pr#18439](#), Nathan Johnson)
- rgw: failed CompleteMultipartUpload request does not release lock ([issue#21596](#), [pr#18430](#), Matt Benjamin)
- rgw\_file: set s->obj\_size from bytes\_written ([issue#21940](#), [pr#18599](#), Matt Benjamin)
- rgw: fix a bug about inconsistent unit of comparison ([issue#21590](#), [pr#18438](#), gaosibei)

- rgw: fix bilog entries on multipart complete ([issue#21772](#), [pr#18334](#), Casey Bodley)
- rgw: fix error handling in ListBucketIndexesCR ([issue#21735](#), [pr#18591](#), Casey Bodley)
- rgw: fix refcnt issues ([issue#21819](#), [pr#18539](#), baixueyu)
- rgw: lc process only schedule the first item of lc objects ([issue#21022](#), [pr#17859](#), Shasha Lu)
- rgw: list bucket which enable versioning get wrong result when user marker ([issue#21500](#), [pr#18569](#), yuliyang)
- rgw: list\_objects() honors end\_marker regardless of namespace ([issue#18977](#), [pr#17832](#), Radoslaw Zarzynski)
- rgw: Multipart upload may double the quota ([issue#21586](#), [pr#18435](#), Sibei Gao)
- rgw: multisite: Get bucket location which is located in another zonegroup, will return 301 Moved Permanently ([issue#21125](#), [pr#17857](#), Shasha Lu)
- rgw: multisite: race between sync of bucket and bucket instance metadata ([issue#21990](#), [pr#18767](#), Casey Bodley)
- rgw: policy checks missing from Get/SetRequestPayment operations ([issue#21389](#), [pr#18440](#), Adam C. Emerson)
- rgw: radosgw-admin usage show loops indefinitely ([issue#21196](#), [pr#18437](#), Mark Kogan)
- rgw: rgw\_file: explicit NFSv3 open() emulation ([issue#21854](#), [pr#18446](#), Matt Benjamin)
- rgw: rgw\_file: fix write error when the write offset overlaps ([issue#21455](#), [pr#18004](#), Yao Zongyou)
- rgw: rgw file write error ([issue#21455](#), [pr#18433](#), Yao Zongyou)
- rgw: s3:GetBucketCORS/s3:PutBucketCORS policy fails with 403 ([issue#21578](#), [pr#18444](#), Adam C. Emerson)
- rgw: s3:GetBucketLocation bucket policy fails with 403 ([issue#21582](#), [pr#18443](#), Adam C. Emerson)
- rgw: s3:GetBucketWebsite/PutBucketWebsite fails with 403 ([issue#21597](#), [pr#18445](#), Adam C. Emerson)
- rgw: setxattrs call leads to different mtimes for bucket index and object ([issue#21200](#), [pr#17856](#), Abhishek Lekshmanan)
- rgw: stop/join TokenCache revoke thread only if started ([issue#21666](#), [pr#18138](#),

Karol Mroz)

- rgw: string\_view instance points to expired memory in PrefixableSignatureHelper ([issue#21085](#), [pr#17858](#), Radoslaw Zarzynski)
- rgw: user creation can overwrite existing user even if different uid is given ([issue#21685](#), [pr#18436](#), Casey Bodley)
- rgw: We can't get torrents if objects are encrypted using SSE-C ([issue#21720](#), [pr#18431](#), Zhang Shaowen)
- rgw: wrong error message is returned when putting container with a name that is too long ([issue#17938](#), [issue#21169](#), [issue#17935](#), [issue#17934](#), [issue#17936](#), [pr#17811](#), Radoslaw Zarzynski)
- rgw: zone compression type is not validated ([issue#21775](#), [pr#18434](#), Casey Bodley)
- tools: ceph-disk create deprecation warnings ([issue#22154](#), [pr#18989](#), Alfredo Deza)
- tools: ceph-disk: fix '-runtime' omission for ceph-osd service ([issue#21498](#), [pr#17914](#), Carl Xiong)
- tools: ceph-disk flake8 test fails on very old, and very new, versions of flake8 ([issue#22207](#), [pr#19152](#), Nathan Cutler)
- tools: ceph-disk: retry on OSError ([issue#21728](#), [pr#18189](#), Kefu Chai)
- tools: ceph-disk: unlocks dmCRYPTED partitions when activating them ([issue#20488](#), [pr#18625](#), Kefu Chai, Felix Winterhalter)
- tools: ceph-kvstore-tool does not call bluestore's umount when exit ([issue#21625](#), [pr#18751](#), Chang Liu)
- tools: ceph\_monstore\_tool: rebuild initial mgrmap also ([issue#22266](#), [pr#19240](#), Kefu Chai)
- tools: ceph-objectstore-tool and ceph-bluestore-tool: backports from master ([issue#21272](#), [pr#17896](#), Sage Weil, David Zafman)
- tools: ceph\_volume\_client: add get, put, and delete object interfaces ([issue#21601](#), [pr#18037](#), Ramana Raja)
- tools: cli/crushtools/build.t sometimes fails in jenkins' make check run ([issue#21758](#), [pr#18398](#), Kefu Chai, Sage Weil)

## v12.2.1 Luminous

---

This is the first bugfix release of Luminous v12.2.x long term stable release series. It contains a range of bug fixes and a few features across CephFS, RBD & RGW. We

recommend all the users of 12.2.x series update.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- Dynamic resharding is now enabled by default for RGW, RGW will now automatically reshards there bucket index once the index grows beyond `rgw_max_objs_per_shard`
- Limiting MDS cache via a memory limit is now supported using the new `mds_cache_memory_limit` config option (1GB by default). A cache reservation can also be specified using `mds_cache_reservation` as a percentage of the limit (5% by default). Limits by inode count are still supported using `mds_cache_size`. Setting `mds_cache_size` to 0 (the default) disables the inode limit.
- The maximum number of PGs per OSD before the monitor issues a warning has been reduced from 300 to 200 PGs. 200 is still twice the generally recommended target of 100 PGs per OSD. This limit can be adjusted via the `mon_max_pg_per_osd` option on the monitors. The older `mon_pg_warn_max_per_osd` option has been removed.
- Creating pools or adjusting `pg_num` will now fail if the change would make the number of PGs per OSD exceed the configured `mon_max_pg_per_osd` limit. The option can be adjusted if it is really necessary to create a pool with more PGs.
- There was a bug in the PG mapping behavior of the new *upmap* feature. If you made use of this feature (e.g., via the `ceph osd pg-upmap-items` command), we recommend that all mappings be removed (via the `ceph osd rm-pg-upmap-items` command) before upgrading to this point release.
- A stall in BlueStore IO submission that was affecting many users has been resolved.

## Other Notable Changes

---

- bluestore: asyn cdeferred\_try\_submit deadlock ([issue#21207](#), [pr#17494](#), Sage Weil)
- bluestore: fix deferred write deadlock, aio short return handling ([issue#21171](#), [pr#17601](#), Sage Weil)
- bluestore: os/bluestore/BlueFS.cc: 1255: FAILED assert(!log\_file->fnodes.empty()) ([issue#21250](#), [pr#17562](#), Sage Weil)
- build/ops: ceph-fuse RPM should require fusermount ([issue#21057](#), [pr#17470](#), Ken Dreyer)

- build/ops: RHEL 7.3 Selinux denials at OSD start ([issue#19200](#), [pr#17468](#), Boris Ranto)
- build/ops: rocksdb,cmake: build portable binaries ([issue#20529](#), [pr#17745](#), Kefu Chai)
- cephfs: client/mds has wrong check to clear S\_ISGID on chown ([issue#21004](#), [pr#17471](#), Patrick Donnelly)
- cephfs: get\_quota\_root sends lookupname op for every buffered write ([issue#20945](#), [pr#17473](#), Dan van der Ster)
- cephfs: MDCache::try\_subtree\_merge() may print N^2 lines of debug message ([issue#21221](#), [pr#17712](#), Patrick Donnelly)
- cephfs: MDS rank add/remove log messages say wrong number of ranks ([issue#21421](#), [pr#17887](#), John Spray)
- cephfs: MDS: standby-replay mds should avoid initiating subtree export ([issue#21378](#), [issue#21222](#), [pr#17714](#), "Yan, Zheng", Jianyu Li)
- cephfs: the standbys are not updated via ceph tell mds.\* command ([issue#21230](#), [pr#17565](#), Kefu Chai)
- common: adding line break at end of some cli results ([issue#21019](#), [pr#17467](#), songweibin)
- core: [cls] metadata\_list API function does not honor max\_return parameter ([issue#21247](#), [pr#17558](#), Jason Dillaman)
- core: incorrect erasure-code space in command ceph df ([issue#21243](#), [pr#17724](#), liuchang0812)
- core: interval\_set: optimize intersect\_of insert operations ([issue#21229](#), [pr#17487](#), Zac Medico)
- core: osd crush rule rename not idempotent ([issue#21162](#), [pr#17481](#), xie xingguo)
- core: osd/PGLog: write only changed dup entries ([issue#21026](#), [pr#17378](#), Josh Durgin)
- doc: doc/rbd: iSCSI Gateway Documentation ([issue#20437](#), [pr#17381](#), Aron Gunn, Jason Dillaman)
- mds: fix 'dirfrag end' check in Server::handle\_client\_readdir ([issue#21070](#), [pr#17686](#), "Yan, Zheng")
- mds: support limiting cache by memory ([issue#20594](#), [pr#17711](#), "Yan, Zheng", Patrick Donnelly)
- mgr: 500 error when attempting to view filesystem data ([issue#20692](#), [pr#17477](#),

John Spray)

- mgr: ceph mgr versions shows active mgr as Unknown ([issue#21260](#), [pr#17635](#), John Spray)
- mgr: Crash in MonCommandCompletion ([issue#21157](#), [pr#17483](#), John Spray)
- mon: mon/OSDMonitor: deleting pool while pgs are being created leads to assert(p != pools.end) in update\_creating\_pgs() ([issue#21309](#), [pr#17634](#), Joao Eduardo Luis)
- mon: OSDMonitor: osd pool application get support ([issue#20976](#), [pr#17472](#), xie xingguo)
- mon: rate limit on health check update logging ([issue#20888](#), [pr#17500](#), John Spray)
- osd: build\_initial\_pg\_history doesn't update up/acting/etc ([issue#21203](#), [pr#17496](#), w11979, Sage Weil)
- osd: osd/PG: discard msgs from down peers ([issue#19605](#), [pr#17501](#), Kefu Chai)
- osd/PrimaryLogPG: request osdmap update in the right block ([issue#21428](#), [pr#17829](#), Josh Durgin)
- osd: PrimaryLogPG: sparse read won't trigger repair correctly ([issue#21123](#), [pr#17475](#), xie xingguo)
- osd: request new map from PG when needed ([issue#21428](#), [pr#17796](#), Josh Durgin)
- osd: Revert "osd/OSDMap: allow bidirectional swap of pg-upmap-items" ([issue#21410](#), [pr#17812](#), Sage Weil)
- osd: subscribe to new osdmap while waiting\_for\_healthy ([issue#21121](#), [pr#17498](#), Sage Weil)
- osd: update info only if new\_interval ([issue#21203](#), [pr#17622](#), Kefu Chai)
- pybind: dashboard usage graph getting bigger and bigger ([issue#20746](#), [pr#17486](#), Yixing Yan)
- rbd: image-meta list does not return all entries ([issue#21179](#), [pr#17561](#), Jason Dillaman)
- rbd: some generic options can not be passed by rbd-nbd ([issue#20426](#), [pr#17557](#), Pan Liu)
- rbd: switch to new config option getter methods ([issue#20737](#), [pr#17464](#), Jason Dillaman)
- rbd: TestMirroringWatcher.ModeUpdated: periodic failure due to injected message failures ([issue#21029](#), [pr#17465](#), Jason Dillaman)

- rgw: bucket index sporadically reshards to 65521 shards ([issue#20934](#), [pr#17476](#), Aleksei Gutikov)
- rgw: bytes\_send and bytes\_recv in the msg of usage show returning is 0 in master branch ([issue#19870](#), [pr#17444](#), Marcus Watts)
- rgw: data encryption sometimes fails to follow AWS settings ([issue#21349](#), [pr#17642](#), hechuang)
- rgw: memory leak in MetadataHandlers ([issue#21214](#), [pr#17570](#), Luo Kexue, Jos Collin)
- rgw: multisite: objects encrypted with SSE-KMS are stored unencrypted in target zone ([issue#20668](#), [issue#20671](#), [pr#17446](#), Casey Bodley)
- rgw: need to stream metadata full sync init ([issue#18079](#), [pr#17448](#), Yehuda Sadeh)
- rgw: object copied from remote src acl permission become full-control issue ([issue#20658](#), [pr#17478](#), Enming Zhang)
- rgw: put lifecycle configuration fails if Prefix is not set ([issue#19587](#), [issue#20872](#), [pr#17479](#), Shasha Lu, Abhishek Lekshmanan)
- rgw: rgw\_file: incorrect lane lock behavior in evict\_block() ([issue#21141](#), [pr#17485](#), Matt Benjamin)
- rgw: send data-log list infinitely ([issue#20951](#), [pr#17445](#), fang.yuxiang)
- rgw: shadow objects are sometimes not removed ([issue#20234](#), [pr#17555](#), Yehuda Sadeh)
- rgw: usage of -inconsistent-index should require user confirmation and print a warning ([issue#20777](#), [pr#17488](#), Orit Wasserman)
- tools: [cli] rename of non-existent image results in seg fault ([issue#21248](#), [pr#17556](#), Jason Dillaman)

## v12.2.0 Luminous

---

This is the first release of Luminous v12.2.x long term stable release series. There have been major changes since Kraken (v11.2.z) and Jewel (v10.2.z), and the upgrade process is non-trivial. Please read these release notes carefully.

## Major Changes from Kraken

---

- General:
  - Ceph now has a simple, [built-in web-based dashboard](#) for monitoring cluster status.

- RADOS:

- BlueStore:

- The new *BlueStore* backend for *ceph-osd* is now stable and the new default for newly created OSDs. BlueStore manages data stored by each OSD by directly managing the physical HDDs or SSDs without the use of an intervening file system like XFS. This provides greater performance and features. See [Storage Devices](#) and [BlueStore Config Reference](#).
    - BlueStore supports [full data and metadata checksums](#) of all data stored by Ceph.
    - BlueStore supports [inline compression](#) using zlib, snappy, or LZ4. (Ceph also supports zstd for [RGW compression](#) but zstd is not recommended for BlueStore for performance reasons.)

- Erasure coded pools now have [full support for overwrites](#), allowing them to be used with RBD and CephFS.

- *ceph-mgr*:

- There is a new daemon, *ceph-mgr*, which is a required part of any Ceph deployment. Although IO can continue when *ceph-mgr* is down, metrics will not refresh and some metrics-related calls (e.g., `ceph df`) may block. We recommend deploying several instances of *ceph-mgr* for reliability. See the notes on [Upgrading](#) below.

- The *ceph-mgr* daemon includes a [REST-based management API](#). The API is still experimental and somewhat limited but will form the basis for API-based management of Ceph going forward.

- *ceph-mgr* also includes a [Prometheus exporter](#) plugin, which can provide Ceph perfcounters to Prometheus.

- *ceph-mgr* now has a [Zabbix](#) plugin. Using `zabbix_sender` it sends trapper events to a Zabbix server containing high-level information of the Ceph cluster. This makes it easy to monitor a Ceph cluster's status and send out notifications in case of a malfunction.

- The overall *scalability* of the cluster has improved. We have successfully tested clusters with up to 10,000 OSDs.

- Each OSD can now have a [device class](#) associated with it (e.g., `hdd` or `ssd`), allowing CRUSH rules to trivially map data to a subset of devices in the system. Manually writing CRUSH rules or manual editing of the CRUSH is normally not required.

- There is a new [upmap](#) exception mechanism that allows individual PGs to be moved around to achieve a *perfect distribution* (this requires luminous

clients).

- Each OSD now adjusts its default configuration based on whether the backing device is an HDD or SSD. Manual tuning generally not required.
- The prototype [mClock QoS queueing algorithm](#) is now available.
- There is now a *backoff* mechanism that prevents OSDs from being overloaded by requests to objects or PGs that are not currently able to process IO.
- There is a simplified [OSD replacement process](#) that is more robust.
- You can query the supported features and (apparent) releases of all connected daemons and clients with [ceph features](#).
- You can configure the oldest Ceph client version you wish to allow to connect to the cluster via `ceph osd set-require-min-compat-client` and Ceph will prevent you from enabling features that will break compatibility with those clients.
- Several sleep settings, include `osd_recovery_sleep`, `osd_snap_trim_sleep`, and `osd_scrub_sleep` have been reimplemented to work efficiently. (These are used in some cases to work around issues throttling background work.)
- Pools are now expected to be associated with the application using them. Upon completing the upgrade to Luminous, the cluster will attempt to associate existing pools to known applications (i.e. CephFS, RBD, and RGW). In-use pools that are not associated to an application will generate a health warning. Any unassociated pools can be manually associated using the new `ceph osd pool application enable` command. For more details see [associate pool to application](#) in the documentation.
- *RGW*:
  - RGW *metadata search* backed by ElasticSearch now supports end user requests service via RGW itself, and also supports custom metadata fields. A query language a set of RESTful APIs were created for users to be able to search objects by their metadata. New APIs that allow control of custom metadata fields were also added.
  - RGW now supports *dynamic bucket index sharding*. This has to be enabled via the `rgw_dyadic_resharding` configurable. As the number of objects in a bucket grows, RGW will automatically reshuffle the bucket index in response. No user intervention or bucket size capacity planning is required.
  - RGW introduces *server side encryption* of uploaded objects with three options for the management of encryption keys: automatic encryption (only recommended for test setups), customer provided keys similar to Amazon SSE-C specification, and through the use of an external key management service (Openstack Barbican) similar to Amazon SSE-KMS specification. [Encryption](#)

- RGW now has preliminary AWS-like bucket policy API support. For now, policy is a means to express a range of new authorization concepts. In the future it will be the foundation for additional auth capabilities such as STS and group policy. [Bucket Policies](#)
- RGW has consolidated the several metadata index pools via the use of rados namespaces. [Pools](#)
- S3 Object Tagging API has been added; while APIs are supported for GET/PUT/DELETE object tags and in PUT object API, there is no support for tags on Policies & Lifecycle yet
- RGW multisite now supports for enabling or disabling sync at a bucket level.
- *RBD:*
  - RBD now has full, stable support for *erasure coded pools* via the new `--data-pool` option to `rbd create`.
  - RBD mirroring's rbd-mirror daemon is now highly available. We recommend deploying several instances of rbd-mirror for reliability.
  - RBD mirroring's rbd-mirror daemon should utilize unique Ceph user IDs per instance to support the new mirroring dashboard.
  - The default 'rbd' pool is no longer created automatically during cluster creation. Additionally, the name of the default pool used by the rbd CLI when no pool is specified can be overridden via a new `rbd default pool = <pool name>` configuration option.
  - Initial support for deferred image deletion via new `rbd trash` CLI commands. Images, even ones actively in-use by clones, can be moved to the trash and deleted at a later time.
  - New pool-level `rbd mirror pool promote` and `rbd mirror pool demote` commands to batch promote/demote all mirrored images within a pool.
  - Mirroring now optionally supports a configurable replication delay via the `rbd mirroring replay delay = <seconds>` configuration option.
  - Improved discard handling when the object map feature is enabled.
  - rbd CLI `import` and `copy` commands now detect sparse and preserve sparse regions.
  - Images and Snapshots will now include a creation timestamp.
  - Specifying user authorization capabilities for RBD clients has been simplified. The general syntax for using RBD capability profiles is "mon 'profile rbd' osd 'profile rbd[-read-only][ pool={pool-name}[, ...]]'". For more details see [User Management](#) in the documentation.

- *CephFS:*

- *Multiple active MDS daemons* is now considered stable. The number of active MDS servers may be adjusted up or down on an active CephFS file system.
- *CephFS directory fragmentation* is now stable and enabled by default on new filesystems. To enable it on existing filesystems use “ceph fs set <fs\_name> allow\_dirfrags”. Large or very busy directories are sharded and (potentially) distributed across multiple MDS daemons automatically.
- Directory subtrees can be explicitly pinned to specific MDS daemons in cases where the automatic load balancing is not desired or effective.
- Client keys can now be created using the new `ceph fs authorize` command to create keys with access to the given CephFS file system and all of its data pools.
- When running ‘df’ on a CephFS filesystem comprising exactly one data pool, the result now reflects the file storage space used and available in that data pool (fuse client only).

- *Miscellaneous:*

- Release packages are now being built for *Debian Stretch*. Note that QA is limited to CentOS and Ubuntu (*xenial* and *trusty*). The distributions we build for now include:
  - CentOS 7 (x86\_64 and aarch64)
  - Debian 8 Jessie (x86\_64)
  - Debian 9 Stretch (x86\_64)
  - Ubuntu 16.04 Xenial (x86\_64 and aarch64)
  - Ubuntu 14.04 Trusty (x86\_64)
- A first release of Ceph for FreeBSD is available which contains a full set of features, other than Bluestore. It will run everything needed to build a storage cluster. For clients, all access methods are available, albeit CephFS is only accessible through a Fuse implementation. RBD images can be mounted on FreeBSD systems through rbd-ggate. Ceph versions are released through the regular FreeBSD ports and packages system. The most current version is available as: net/ceph-devel. Once Luminous goes into official release, this version will be available as net/ceph. Future development releases will be available via net/ceph-devel. More details about this port are in: [README.FreeBSD](#)
- *CLI changes:*
  - The `ceph -s` or `ceph status` command has a fresh look.

- `ceph mgr metadata` will dump metadata associated with each mgr daemon.
- `ceph versions` or `ceph {osd,mds,mon,mgr} versions` summarize versions of running daemons.
- `ceph {osd,mds,mon,mgr} count-metadata <property>` similarly tabulates any other daemon metadata visible via the `ceph {osd,mds,mon,mgr} metadata` commands.
- `ceph features` summarizes features and releases of connected clients and daemons.
- `ceph osd require-osd-release <release>` replaces the old `require_RELEASE_osds` flags.
- `ceph osd pg-upmap`, `ceph osd rm-pg-upmap`, `ceph osd pg-upmap-items`, `ceph osd rm-pg-upmap-items` can explicitly manage upmap items (see [Using the pg-upmap](#)).
- `ceph osd getcrushmap` returns a crush map version number on stderr, and `ceph osd setcrushmap [version]` will only inject an updated crush map if the version matches. This allows crush maps to be updated offline and then reinjected into the cluster without fear of clobbering racing changes (e.g., by newly added osds or changes by other administrators).
- `ceph osd create` has been replaced by `ceph osd new`. This should be hidden from most users by user-facing tools like ceph-disk.
- `ceph osd destroy` will mark an OSD destroyed and remove its cephx and lockbox keys. However, the OSD id and CRUSH map entry will remain in place, allowing the id to be reused by a replacement device with minimal data rebalancing.
- `ceph osd purge` will remove all traces of an OSD from the cluster, including its cephx encryption keys, dm-crypt lockbox keys, OSD id, and crush map entry.
- `ceph osd ls-tree <name>` will output a list of OSD ids under the given CRUSH name (like a host or rack name). This is useful for applying changes to entire subtrees. For example, `ceph osd down `ceph osd ls-tree rack1``.
- `ceph osd {add,rm}-{noout,noin,nodown,noup}` allow the noout, noin, nodown, and noup flags to be applied to specific OSDs.
- `ceph osd safe-to-destroy <osd(s)>` will report whether it is safe to remove or destroy OSD(s) without reducing data durability or redundancy.
- `ceph osd ok-to-stop <osd(s)>` will report whether it is okay to stop OSD(s) without immediately compromising availability (i.e., all PGs should remain active but may be degraded).
- `ceph log last [n]` will output the last *n* lines of the cluster log.

- `ceph mgr dump` will dump the MgrMap, including the currently active ceph-mgr daemon and any standbys.
- `ceph mgr module ls` will list active ceph-mgr modules.
- `ceph mgr module {enable,disable} <name>` will enable or disable the named mgr module. The module must be present in the configured `mgr_module_path` on the host(s) where `ceph-mgr` is running.
- `ceph osd crush ls <node>` will list items (OSDs or other CRUSH nodes) directly beneath a given CRUSH node.
- `ceph osd crush swap-bucket <src> <dest>` will swap the contents of two CRUSH buckets in the hierarchy while preserving the buckets' ids. This allows an entire subtree of devices to be replaced (e.g., to replace an entire host of FileStore OSDs with newly-imaged BlueStore OSDs) without disrupting the distribution of data across neighboring devices.
- `ceph osd set-require-min-compat-client <release>` configures the oldest client release the cluster is required to support. Other changes, like CRUSH tunables, will fail with an error if they would violate this setting. Changing this setting also fails if clients older than the specified release are currently connected to the cluster.
- `ceph config-key dump` dumps config-key entries and their contents. (The existing `ceph config-key list` only dumps the key names, not the values.)
- `ceph config-key list` is deprecated in favor of `ceph config-key ls`.
- `ceph config-key put` is deprecated in favor of `ceph config-key set`.
- `ceph auth list` is deprecated in favor of `ceph auth ls`.
- `ceph osd crush rule list` is deprecated in favor of `ceph osd crush rule ls`.
- `ceph osd set-{full,nearfull,backfillfull}-ratio` sets the cluster-wide ratio for various full thresholds (when the cluster refuses IO, when the cluster warns about being close to full, when an OSD will defer rebalancing a PG to itself, respectively).
- `ceph osd reweightn` will specify the reweight values for multiple OSDs in a single command. This is equivalent to a series of `ceph osd reweight` commands.
- `ceph osd crush {set,rm}-device-class` manage the new CRUSH *device class* feature. Note that manually creating or deleting a device class name is generally not necessary as it will be smart enough to be self-managed. `ceph osd crush class ls` and `ceph osd crush class ls-osd` will output all existing device classes and a list of OSD ids under the given device class respectively.

- `ceph osd crush rule create-replicated` replaces the old `ceph osd crush rule create-simple` command to create a CRUSH rule for a replicated pool. Notably it takes a class argument for the *device class* the rule should target (e.g., `ssd` or `hdd`).
- `ceph mon feature ls` will list monitor features recorded in the MonMap. `ceph mon feature set` will set an optional feature (none of these exist yet).
- `ceph tell <daemon> help` will now return a usage summary.
- `ceph fs authorize` creates a new client key with caps automatically set to access the given CephFS file system.
- The `ceph health` structured output (JSON or XML) no longer contains ‘timechecks’ section describing the time sync status. This information is now available via the ‘ceph time-sync-status’ command.
- Certain extra fields in the `ceph health` structured output that used to appear if the mons were low on disk space (which duplicated the information in the normal health warning messages) are now gone.
- The `ceph -w` output no longer contains audit log entries by default. Add a `--watch-channel=audit` or `--watch-channel=*` to see them.
- New “ceph -w” behavior - the “ceph -w” output no longer contains I/O rates, available space, pg info, etc. because these are no longer logged to the central log (which is what `ceph -w` shows). The same information can be obtained by running `ceph pg stat`; alternatively, I/O rates per pool can be determined using `ceph osd pool stats`. Although these commands do not self-update like `ceph -w` did, they do have the ability to return formatted output by providing a `--format=<format>` option.
- Added new commands `pg force-recovery` and `pg-force-backfill`. Use them to boost recovery or backfill priority of specified pgs, so they’re recovered/backfilled before any other. Note that these commands don’t interrupt ongoing recovery/backfill, but merely queue specified pgs before others so they’re recovered/backfilled as soon as possible. New commands `pg cancel-force-recovery` and `pg cancel-force-backfill` restore default recovery/backfill priority of previously forced pgs.

# Major Changes from Jewel

- *RADOS*:
  - We now default to the AsyncMessenger (`ms type = async`) instead of the legacy SimpleMessenger. The most noticeable difference is that we now use a fixed sized thread pool for network connections (instead of two threads per socket with SimpleMessenger).
  - Some OSD failures are now detected almost immediately, whereas previously the heartbeat timeout (which defaults to 20 seconds) had to expire. This prevents IO from blocking for an extended period for failures where the host remains up but the ceph-osd process is no longer running.
  - The size of encoded OSDMaps has been reduced.
  - The OSDs now quiesce scrubbing when recovery or rebalancing is in progress.
- *RGW*:
  - RGW now supports the S3 multipart object copy-part API.
  - It is possible now to reshuffle an existing bucket offline. Offline bucket reshuffling currently requires that all IO (especially writes) to the specific bucket is quiesced. (For automatic online reshuffling, see the new feature in Luminous above.)
  - RGW now supports data compression for objects.
  - Civetweb version has been upgraded to 1.8
  - The Swift static website API is now supported (S3 support has been added previously).
  - S3 bucket lifecycle API has been added. Note that currently it only supports object expiration.
  - Support for custom search filters has been added to the LDAP auth implementation.
  - Support for NFS version 3 has been added to the RGW NFS gateway.
  - A Python binding has been created for librgw.
- *RBD*:
  - The rbd-mirror daemon now supports replicating dynamic image feature updates and image metadata key/value pairs from the primary image to the non-primary image.

- The number of image snapshots can be optionally restricted to a configurable maximum.
- The rbd Python API now supports asynchronous IO operations.
- *CephFS*:
  - libcephfs function definitions have been changed to enable proper uid/gid control. The library version has been increased to reflect the interface change.
  - Standby replay MDS daemons now consume less memory on workloads doing deletions.
  - Scrub now repairs backtrace, and populates damage ls with discovered errors.
  - A new pg\_files subcommand to cephfs-data-scan can identify files affected by a damaged or lost RADOS PG.
  - The false-positive “failing to respond to cache pressure” warnings have been fixed.

## Upgrade from Jewel or Kraken

1. Ensure that the `sortbitwise` flag is enabled:

```
1. # ceph osd set sortbitwise
```

2. Make sure your cluster is stable and healthy (no down or recovering OSDs). (Optional, but recommended.)
3. Do not create any new erasure-code pools while upgrading the monitors.
4. You can monitor the progress of your upgrade at each stage with the `ceph versions` command, which will tell you what ceph version is running for each type of daemon.
5. Set the `noout` flag for the duration of the upgrade. (Optional but recommended.):

```
1. # ceph osd set noout
```

6. Verify that all RBD client users have sufficient caps to blacklist other client users. RBD client users with only `"allow r"` monitor caps should be updated as follows:

```
# ceph auth caps client.<ID> mon 'allow r, allow command "osd blacklist"' osd '<existing OSD caps for
1. user>'
```

- Upgrade monitors by installing the new packages and restarting the monitor daemons. Note that, unlike prior releases, the ceph-mon daemons *must* be upgraded first:

```
1. # systemctl restart ceph-mon.target
```

Verify the monitor upgrade is complete once all monitors are up by looking for the `luminous` feature string in the mon map. For example:

```
1. # ceph mon feature ls
```

should include luminous under persistent features:

```
1. on current monmap (epoch NNN)
2.   persistent: [kraken, luminous]
3.   required: [kraken, luminous]
```

- Add or restart `ceph-mgr` daemons. If you are upgrading from kraken, upgrade packages and restart ceph-mgr daemons with:

```
1. # systemctl restart ceph-mgr.target
```

If you are upgrading from kraken, you may already have ceph-mgr daemons deployed. If not, or if you are upgrading from jewel, you can deploy new daemons with tools like ceph-deploy or ceph-ansible. For example:

```
1. # ceph-deploy mgr create HOST
```

Verify the ceph-mgr daemons are running by checking `ceph -s` :

```
1. # ceph -s
2.
3. ...
4. services:
5.   mon: 3 daemons, quorum foo,bar,baz
6.   mgr: foo(active), standbys: bar, baz
7. ...
```

- Upgrade all OSDs by installing the new packages and restarting the ceph-osd daemons on all hosts:

```
1. # systemctl restart ceph-osd.target
```

You can monitor the progress of the OSD upgrades with the new `ceph versions` or `ceph osd versions` command:

```
1. # ceph osd versions
2. {
3.     "ceph version 12.2.0 (...) luminous (stable)": 12,
4.     "ceph version 10.2.6 (...)": 3,
5. }
```

10. Upgrade all CephFS daemons by upgrading packages and restarting daemons on all hosts:

```
1. # systemctl restart ceph-mds.target
```

11. Upgrade all radosgw daemons by upgrading packages and restarting daemons on all hosts:

```
1. # systemctl restart radosgw.target
```

12. Complete the upgrade by disallowing pre-luminous OSDs and enabling all new Luminous-only functionality:

```
1. # ceph osd require-osd-release luminous
```

If you set `noout` at the beginning, be sure to clear it with:

```
1. # ceph osd unset noout
```

13. Verify the cluster is healthy with `ceph health`.

## Upgrading from pre-Jewel releases (like Hammer)

You *must* first upgrade to Jewel (10.2.z) before attempting an upgrade to Luminous.

# Upgrade compatibility notes, Jewel to Kraken

These changes occurred between the Jewel and Kraken releases and will affect upgrades from Jewel to Luminous.

- The `osd crush location` config option is no longer supported. Please update your `ceph.conf` to use the `crush location` option instead. Be sure to update your config file to avoid any movement of OSDs from your customized location back to the default one.
- The OSDs now avoid starting new scrubs while recovery is in progress. To revert to the old behavior (and do not let recovery activity affect the scrub scheduling) you can set the following option:

```
1. osd scrub during recovery = true
```

- The list of monitor hosts/addresses for building the monmap can now be obtained from DNS SRV records. The service name used in when querying the DNS is defined in the “`mon_dns_srv_name`” config option, which defaults to “`ceph-mon`”.
- The ‘`osd class load list`’ config option is a list of object class names that the OSD is permitted to load (or ‘`*`’ for all classes). By default it contains all existing in-tree classes for backwards compatibility.
- The ‘`osd class default list`’ config option is a list of object class names (or ‘`*`’ for all classes) that clients may invoke having only the ‘`*`’, ‘`x`’, ‘`class-read`’, or ‘`class-write`’ capabilities. By default it contains all existing in-tree classes for backwards compatibility. Invoking classes not listed in ‘`osd class default list`’ requires a capability naming the class (e.g. ‘`allow class foo`’).
- The ‘`rgw rest getusage op compat`’ config option allows you to dump (or not dump) the description of user stats in the S3 GetUsage API. This option defaults to false. If the value is true, the response data for GetUsage looks like:

```
1. "stats": {
2.     "TotalBytes": 516,
3.     "TotalBytesRounded": 1024,
4.     "TotalEntries": 1
5. }
```

If the value is false, the response for GetUsage looks as it did before:

```
1. {
2.     516,
3.     1024,
4.     1
5. }
```

- The ‘osd out ...’ and ‘osd in ...’ commands now preserve the OSD weight. That is, after marking an OSD out and then in, the weight will be the same as before (instead of being reset to 1.0). Previously the mons would only preserve the weight if the mon automatically marked an OSD out and then in, but not when an admin did so explicitly.
- The ‘ceph osd perf’ command will display ‘commit\_latency(ms)’ and ‘apply\_latency(ms)’. Previously, the names of these two columns were ‘fs\_commit\_latency(ms)’ and ‘fs\_apply\_latency(ms)’. We removed the prefix ‘fs\_’ because the values are not filestore-specific.
- Monitors will no longer allow pools to be removed by default. The setting `mon_allow_pool_delete` has to be set to true (defaults to false) before they allow pools to be removed. This is an additional safeguard against pools being removed by accident.
- If you have manually specified the monitor user `rocksdb` via the `mon keyvaluedb = rocksdb` option, you will need to manually add a file to the mon data directory to preserve this option:

```
1. echo rocksdb > /var/lib/ceph/mon/ceph-`hostname`/kv_backend
```

New monitors will now use `rocksdb` by default, but if that file is not present, existing monitors will use `leveldb`. The `mon keyvaluedb` option now only affects the backend chosen when a monitor is created.

- The ‘osd crush initial weight’ option allows you to specify a CRUSH weight for a newly added OSD. Previously a value of 0 (the default) meant that we should use the size of the OSD’s store to weight the new OSD. Now, a value of 0 means it should have a weight of 0, and a negative value (the new default) means we should automatically weight the OSD based on its size. If your configuration file explicitly specifies a value of 0 for this option you will need to change it to a negative value (e.g., -1) to preserve the current behavior.
- The static libraries are no longer included by the debian development packages (`lib*-dev`) as it is not required per debian packaging policy. The shared (.so) versions are packaged as before.
- The libtool pseudo-libraries (.la files) are no longer included by the debian development packages (`lib*-dev`) as they are not required per <https://wiki.debian.org/ReleaseGoals/LAFileRemoval> and <https://www.debian.org/doc/manuals/maint-guide/advanced.en.html>.
- The jerasure and shec plugins can now detect SIMD instruction at runtime and no longer need to be explicitly configured for different processors. The following plugins are now deprecated: `jerasure_generic`, `jerasure_sse3`, `jerasure_sse4`, `jerasure_neon`, `shec_generic`, `shec_sse3`, `shec_sse4`, and `shec_neon`. If you use any

of these plugins directly you will see a warning in the mon log file. Please switch to using just ‘jerasure’ or ‘shec’.

- The librados omap get\_keys and get\_vals operations include a start key and a limit on the number of keys to return. The OSD now imposes a configurable limit on the number of keys and number of total bytes it will respond with, which means that a librados user might get fewer keys than they asked for. This is necessary to prevent careless users from requesting an unreasonable amount of data from the cluster in a single operation. The new limits are configured with `osd_max_omap_entries_per_request`, defaulting to 131,072, and `osd_max_omap_bytes_per_request`, defaulting to 4MB.
- Calculation of recovery priorities has been updated. This could lead to unintuitive recovery prioritization during cluster upgrade. In case of such recovery, OSDs in old version would operate on different priority ranges than new ones. Once upgraded, cluster will operate on consistent values.

## Upgrade compatibility notes, Kraken to Luminous

---

- The configuration option `osd pool erasure code stripe width` has been replaced by `osd pool erasure code stripe unit`, and given the ability to be overridden by the erasure code profile setting `stripe_unit`. For more details see [Erasure code profiles](#).
- rbd and cephfs can use erasure coding with bluestore. This may be enabled by setting `allow_ec_overwrites` to `true` for a pool. Since this relies on bluestore’s checksumming to do deep scrubbing, enabling this on a pool stored on filestore is not allowed.
- The `rados df` JSON output now prints numeric values as numbers instead of strings.
- The `mon_osd_max_op_age` option has been renamed to `mon_osd_warn_op_age` (default: 32 seconds), to indicate we generate a warning at this age. There is also a new `mon_osd_err_op_age_ratio` that is expressed as a multiple of `mon_osd_warn_op_age` (default: 128, for roughly 60 minutes) to control when an error is generated.
- The default maximum size for a single RADOS object has been reduced from 100GB to 128MB. The 100GB limit was completely impractical in practice while the 128MB limit is a bit high but not unreasonable. If you have an application written directly to librados that is using objects larger than 128MB you may need to adjust `osd_max_object_size`.
- The semantics of the `rados ls` and librados object listing operations have always been a bit confusing in that “whiteout” objects (which logically don’t exist and will return ENOENT if you try to access them) are included in the results. Previously whiteouts only occurred in cache tier pools. In luminous, logically deleted but snapshotted objects now result in a whiteout object, and as a result

they will appear in `rados ls` results, even though trying to read such an object will result in ENOENT. The `rados listsnaps` operation can be used in such a case to enumerate which snapshots are present. This may seem a bit strange, but is less strange than having a deleted-but-snapshotned object not appear at all and be completely hidden from librados's ability to enumerate objects. Future versions of Ceph will likely include an alternative object enumeration interface that makes it more natural and efficient to enumerate all objects along with their snapshot and clone metadata.

- The deprecated `crush_ruleset` property has finally been removed; please use `crush_rule` instead for the `osd pool get ...` and `osd pool set ...` commands.
- The `osd pool default crush replicated ruleset` option has been removed and replaced by the `osd pool default crush rule` option. By default it is -1, which means the mon will pick the first type replicated rule in the CRUSH map for replicated pools. Erasure coded pools have rules that are automatically created for them if they are not specified at pool creation time.
- We no longer test the FileStore ceph-osd backend in combination with btrfs. We recommend against using btrfs. If you are using btrfs-based OSDs and want to upgrade to luminous you will need to add the following to your ceph.conf:

```
1. enable experimental unrecoverable data corrupting features = btrfs
```

The code is mature and unlikely to change, but we are only continuing to test the Jewel stable branch against btrfs. We recommend moving these OSDs to FileStore with XFS or BlueStore.

- The `ruleset-*` properties for the erasure code profiles have been renamed to `crush-*` to move away from the obsolete 'ruleset' term and to be more clear about their purpose. There is also a new optional `crush-device-class` property to specify a CRUSH device class to use for the erasure coded pool. Existing erasure code profiles will be converted automatically when upgrade completes (when the `ceph osd require-osd-release luminous` command is run) but any provisioning tools that create erasure coded pools may need to be updated.
- The structure of the XML output for `osd crush tree` has changed slightly to better match the `osd tree` output. The top level structure is now `nodes` instead of `crush_map_roots`.
- When assigning a network to the public network and not to the cluster network the network specification of the public network will be used for the cluster network as well. In older versions this would lead to cluster services being bound to `0.0.0.0:<port>`, thus making the cluster service even more publicly available than the public services. When only specifying a cluster network it will still result in the public services binding to `0.0.0.0`.
- In previous versions, if a client sent an op to the wrong OSD, the OSD would

reply with ENXIO. The rationale here is that the client or OSD is clearly buggy and we want to surface the error as clearly as possible. We now only send the ENXIO reply if the `osd_enxio_on_misdirected_op` option is enabled (it's off by default). This means that a VM using librbd that previously would have gotten an EIO and gone read-only will now see a blocked/hung IO instead.

- The “journaler allow split entries” config setting has been removed.
- The ‘`mon_warn_osd_usage_min_max_delta`’ config option has been removed and the associated health warning has been disabled because it does not address clusters undergoing recovery or CRUSH rules that do not target all devices in the cluster.
- Added new configuration “public bind addr” to support dynamic environments like Kubernetes. When set the Ceph MON daemon could bind locally to an IP address and advertise a different IP address `public_addr` on the network.
- The crush `choose_args` encoding has been changed to make it architecture-independent. If you deployed Luminous dev releases or 12.1.0 rc release and made use of the CRUSH choose\_args feature, you need to remove all `choose_args` mappings from your CRUSH map before starting the upgrade.
- *librados:*
  - Some variants of the `omap_get_keys` and `omap_get_vals` librados functions have been deprecated in favor of `omap_get_vals2` and `omap_get_keys2`. The new methods include an output argument indicating whether there are additional keys left to fetch. Previously this had to be inferred from the requested key count vs the number of keys returned, but this breaks with new OSD-side limits on the number of keys or bytes that can be returned by a single omap request. These limits were introduced by kraken but are effectively disabled by default (by setting a very large limit of 1 GB) because users of the newly deprecated interface cannot tell whether they should fetch more keys or not. In the case of the standalone calls in the C++ interface (`IoCtx::get_omap_{keys,vals}`), librados has been updated to loop on the client side to provide a correct result via multiple calls to the OSD. In the case of the methods used for building multi-operation transactions, however, client-side looping is not practical, and the methods have been deprecated. Note that use of either the `IoCtx` methods on older librados versions or the deprecated methods on any version of librados will lead to incomplete results if/when the new OSD limits are enabled.
  - The original librados `rados_objects_list_open` (C) and `objects_begin` (C++) object listing API, deprecated in Hammer, has finally been removed. Users of this interface must update their software to use either the `rados_nobjects_list_open` (C) and `nobjects_begin` (C++) API or the new `rados_object_list_begin` (C) and `object_list_begin` (C++) API before updating the client-side librados library to Luminous. Object enumeration (via any API) with the latest librados version and pre-Hammer OSDs is no longer

supported. Note that no in-tree Ceph services rely on object enumeration via the deprecated APIs, so only external librados users might be affected. The newest (and recommended) rados\_object\_list\_begin (C) and object\_list\_begin (C++) API is only usable on clusters with the SORTBITWISE flag enabled (Jewel and later). (Note that this flag is required to be set before upgrading beyond Jewel.)

- *CephFS*:

- When configuring ceph-fuse mounts in /etc/fstab, a new syntax is available that uses “ceph.<arg>=<val>” in the options column, instead of putting configuration in the device column. The old style syntax still works. See the documentation page “Mount CephFS in your file systems table” for details.
- CephFS clients without the ‘p’ flag in their authentication capability string will no longer be able to set quotas or any layout fields. This flag previously only restricted modification of the pool and namespace fields in layouts.
- CephFS will generate a health warning if you have fewer standby daemons than it thinks you wanted. By default this will be 1 if you ever had a standby, and 0 if you did not. You can customize this using `ceph fs set <fs> standby_count_wanted <number>`. Setting it to zero will effectively disable the health check.
- The “ceph mds tell ...” command has been removed. It is superseded by “ceph tell mds.<id> ...”
- The `apply` mode of cephfs-journal-tool has been removed

## Other Notable Changes

- `async`: Fixed compilation error when enable `-DWITH_DPDK` ([pr#12660](#), Pan Liu)
- `async`: fixed coredump when enable dpdk ([pr#12854](#), Pan Liu)
- `async`: fixed the error "Cause: Cannot create lock on '/var/run/.rte\_c..." ([pr#12860](#), Pan Liu)
- `bluestore`: avoid unnecessary copy with `coll_t` ([pr#12576](#), Yunchuan Wen)
- `bluestore`: `BitAllocator`: delete useless codes ([pr#13619](#), Jie Wang)
- `bluestore`: `bluestore/BlueFS`: pass string as const ref ([pr#16600](#), dingdangzhang)
- `bluestore`: `bluestore, NVMEDEVICE`: Specify the max io completion in conf ([pr#13799](#), optimistyzy)
- `bluestore`: `bluestore/NVMEDEVICE`: update SPDK to version 17.03 ([pr#14585](#), optimistyzy)
- `bluestore`: `bluestore, NVMeDevice`: use task' own lock for (random) read ([pr#14094](#), optimistyzy)
- `bluestore, build/ops, performance`: `os/bluestore`: enable SSE-assisted CRC32 calculations in RocksDB ([pr#13741](#), Radoslaw Zarzynski)
- `bluestore`: `ceph-disk`: add `-filestore` argument, default to `-bluestore` ([pr#15437](#), Loic Dachary, Sage Weil)
- `bluestore`: `common/config`: set `rocksdb_cache_size` to `OPT_U64` ([pr#13995](#), liuhongtong)
- `bluestore`: `common/options`: make "`blue{fs, store}_allocator`" `LEVEL_DEV` ([issue#20660](#), [pr#16645](#), Kefu Chai)
- `bluestore, common, performance`: `common/Finisher`: Using `queue(list<context*>)` instead `queue(context*)` ([pr#8942](#), Jianpeng Ma)
- `bluestore, common, performance`: `isa-l`: update `isa-l` to v2.18 ([pr#15895](#), Ganesh Mahalingam, Tushar Gohad)
- `bluestore, core`: `os/bluestore`: fix statfs to not include DB partition in free space ([issue#18599](#), [pr#13140](#), Sage Weil)
- `bluestore, core`: `os/bluestore`: fix warning ([pr#15435](#), Sage Weil)
- `bluestore, core`: `os/bluestore`: improve mempool usage ([pr#15402](#), Sage Weil)
- `bluestore, core`: `os/bluestore`: write "`mkfs_done`" into disk only if we pass `fsck()`

- tests ([pr#15238](#), xie xingguo)
- bluestore,core: osd/OSDMap: should update input param if osd dne ([pr#14863](#), Kefu Chai)
  - bluestore,core: os: remove experimental status for BlueStore ([pr#15177](#), Sage Weil)
  - bluestore: fixed compilation error when enable spdk ([pr#12672](#), Pan Liu)
  - bluestore: include/intarith: templatize ctz/clz/cbits helpers ([pr#14862](#), Kefu Chai)
  - bluestore: luminous: os/bluestore: compensate for bad freelistmanager size/blocks metadata ([issue#21089](#), [pr#17273](#), Sage Weil)
  - bluestore: NVMEDevice: add the spdk core mask check ([pr#14068](#), optimistyzy)
  - bluestore: NVMEDevice: cleanup the logic in data\_buf\_next\_sge ([pr#13056](#), optimistyzy)
  - bluestore: NVMeDevice: fix the core id for rte\_remote\_launch ([pr#13896](#), optimistyzy)
  - bluestore: NVMEDevice: fix bug in data\_buf\_next\_sge ([pr#12812](#), optimistyzy)
  - bluestore: NVMEDevice: minor error for get slave core ([pr#14012](#), Ziye Yang)
  - bluestore: NVMEDevice: optimize sector\_size usage ([pr#12780](#), optimistyzy)
  - bluestore: NVMEDevice: remove unnecessary dpdk header file ([pr#14650](#), optimistyzy)
  - bluestore: NVMEDevice: fix the I/O logic for read ([pr#13971](#), optimistyzy)
  - bluestore: os/bluestore: add a debug option to bypass block device writes for bl... ([pr#12464](#), Igor Fedotov)
  - bluestore: os/bluestore: Add bluestore pextent vector to mempool ([pr#12946](#), Igor Fedotvo, Igor Fedotov)
  - bluestore: os/bluestore: add flush\_store\_cache cmd ([pr#13428](#), xie xingguo)
  - bluestore: os/bluestore: add more perf\_counters to BlueStore ([pr#13274](#), Igor Fedotov)
  - bluestore: os/bluestore: add new garbage collector ([pr#12144](#), Igor Fedotov)
  - bluestore: os/bluestore: add perf variable for throttle info in bluestore ([pr#12583](#), Pan Liu)
  - bluestore: os/bluestore: add “\_” prefix for internal methods ([pr#13409](#), xie xingguo)

- bluestore: os/bluestore: align reclaim size to bluefs\_alloc\_size ([pr#14744](#), Haomai Wang)
- bluestore: os/bluestore/Allocator: drop unused return value in release function ([pr#13913](#), wangzhengyong)
- bluestore: os/bluestore: allow multiple DeferredBatches in flight at once ([issue#20295](#), [pr#16769](#), Nathan Cutler, Sage Weil)
- bluestore: os/bluestore: allow multiple SPDK BlueStore OSD instances ([issue#16966](#), [pr#12604](#), Orlando Moreno)
- bluestore: os/bluestore: assert blob map returns success ([pr#14473](#), shiqi)
- bluestore: os/bluestore: avoid nullptr in bluestore\_extent\_ref\_map\_t::bound\_encode ([pr#14073](#), Sage Weil)
- bluestore: os/bluestore: avoid unnecessary memory copy, use variable reference in BlockDevice::Open ([pr#12942](#), liuchang0812)
- bluestore: os/bluestore: better debug output on unsharing blobs ([issue#20227](#), [pr#15746](#), Sage Weil)
- bluestore: os/bluestore/BitAllocator: fix bug of checking required blocks ([pr#13470](#), wangzhengyong)
- bluestore: os/bluestore/BitMapAllocator: rm unused variable ([pr#13599](#), Jie Wang)
- bluestore: os/bluestore/BitmapFreelistManager: readability improvements ([pr#12719](#), xie xingguo)
- bluestore: os/bluestore/BlockDevice: support pmem device as bluestore backend ([pr#15102](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: clean up log\_writer aios from compaction ([issue#20454](#), [pr#16017](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: clear current log entrys before dump all fnode ([pr#15973](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: fix reclaim\_blocks ([issue#18368](#), [pr#12725](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: Rebuild memcpy for bufferlist::page\_aligned\_app... ([pr#15728](#), Jianpeng Ma, Sage Weil)
- bluestore: os/bluestore/BlueFS: .slow should be compared with dirname ([pr#15595](#), zanglei)
- bluestore: os/bluestore/BlueStore: Avoid double counting state\_kv\_queued\_lat ([pr#16374](#), Jianpeng Ma)

- bluestore: os/bluestore/BlueStore.cc:remove unuse code in \_open\_bdev() ([pr#13553](#), yonghengdexin735)
- bluestore: os/bluestore/BlueStore.cc: remove unused variable ([pr#12703](#), Li Wang)
- bluestore: os/bluestore/BlueStore: no device no symlink ([pr#15721](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueStore: remove unused code ([pr#16522](#), Jianpeng Ma)
- bluestore: os/bluestore: cleanup BitAllocator ([pr#12661](#), xie xingguo)
- bluestore: os/bluestore: cleanup bluestore\_types ([pr#15680](#), xie xingguo)
- bluestore: os/bluestore: clean up flush logic ([pr#14162](#), Jianpeng Ma)
- bluestore: os/bluestore: cleanup, got rid of table reference of 1<<x ([pr#13718](#), Adam Kupczyk)
- bluestore: os/bluestore: clean up Invalid return value judgment ([pr#14219](#), shiqi)
- bluestore: os/bluestore: cleanup min\_alloc\_size; some formatting nits ([pr#15826](#), xie xingguo)
- bluestore: os/bluestore: clear result in BlueRocksEnv::getChildren ([issue#20857](#), [pr#16683](#), liuchang0812)
- bluestore: os/bluestore: clear up redundant size assignment in KerenelDevice ([pr#16121](#), Shasha Lu)
- bluestore: os/bluestore: conditionally load crr option ([pr#12877](#), xie xingguo)
- bluestore: os/bluestore: configure rocksdb cache via bluestore\_cache\_kv\_ratio ([pr#15580](#), Sage Weil)
- bluestore: os/bluestore: default journal media to store media if bluefs is disabled ([pr#16844](#), xie xingguo)
- bluestore: os/bluestore: \_do\_remove: dirty shard individually as each blob is unshared ([issue#20849](#), [pr#16822](#), Sage Weil)
- bluestore: os/blueStore: Failure retry for opening file ([pr#16237](#), Yankun Li)
- bluestore: os/bluestore: fix a bug in small write handling on sharded extents ([pr#13728](#), Igor Fedotov)
- bluestore: os/bluestore: fix Allocator::allocate() int truncation ([issue#18595](#), [pr#13010](#), Sage Weil)
- bluestore: os/bluestore: fix a typo about bleustore ([pr#15357](#), Dongsheng Yang)
- bluestore: os/bluestore: fix BitMapAllocator assert on out-of-bound hint value ([pr#15289](#), Igor Fedotov)

- bluestore: os/bluestore: fix buffers pinned by indefinitely deferred writes ([pr#15398](#), Sage Weil)
- bluestore: os/bluestore: fix bug for calc\_extent\_avg in reshard function ([pr#13931](#), wangzhengyong)
- bluestore: os/bluestore: fix bug in aio\_read() ([pr#13511](#), tangwenjun)
- bluestore: os/bluestore: fix bug in \_open\_alloc() ([pr#13577](#), yonghengdexin735)
- bluestore: os/bluestore: fix bug in \_open\_super\_meta() ([pr#13559](#), Taeksang Kim)
- bluestore: os/bluestore: fix bugs in bluefs and bdev flush ([issue#19250](#), [issue#19251](#), [pr#13911](#), Sage Weil)
- bluestore: os/bluestore: fix coredump in register\_ctrlr() ([pr#13556](#), tangwenjun)
- bluestore: os/bluestore: fix deferred\_aio deadlock ([pr#16051](#), Sage Weil)
- bluestore: os/bluestore: fix deferred write race ([issue#19880](#), [pr#15004](#), Sage Weil)
- bluestore: os/bluestore: fix deferred writes vs collection split race ([issue#19379](#), [pr#14157](#), Sage Weil)
- bluestore: os/bluestore: fix dirty\_range on \_do\_clone\_range ([issue#20810](#), [pr#16738](#), Sage Weil)
- bluestore: os/bluestore: fix false assert in IOContext::aio\_wake ([pr#15268](#), Igor Fedotov)
- bluestore: os/bluestore: fix false asserts in Cache::trim\_all() ([pr#15470](#), xie xingguo)
- bluestore: os/bluestore: fix fsck deferred\_replay ([pr#15295](#), Sage Weil)
- bluestore: os/bluestore: fix min\_alloc\_size at mkfs time ([pr#13192](#), Sage Weil)
- bluestore: os/bluestore: fix narrow osr->flush() race ([pr#14489](#), Sage Weil)
- bluestore: os/bluestore: fix NVMEDevice::open failure if serial number ends with a ... ([pr#12956](#), Hongtong Liu)
- bluestore: os/bluestore: fix OnodeSizeTracking testing ([issue#20498](#), [pr#12684](#), xie xingguo)
- bluestore: os/bluestore: fix perf counters ([pr#13965](#), Sage Weil)
- bluestore: os/bluestore: fix possible out of order shard(offset == 0); add sanity check ([pr#15658](#), xie xingguo)
- bluestore: os/bluestore: fix potential access violation ([pr#15657](#), xie xingguo)

- bluestore: os/bluestore: fix potential assert in cache \_trim method ([pr#13234](#), Igor Fedotov)
- bluestore: os/bluestore: fix reclaim\_blocks and clean up Allocator interface ([issue#18573](#), [pr#12963](#), Sage Weil)
- bluestore: os/bluestore: fix typo(s/trasnaction/transaction/) ([pr#14890](#), xie xingguo)
- bluestore: os/bluestore: fix unsharing blob dirty\_range args ([issue#20227](#), [pr#15766](#), Sage Weil)
- bluestore: os/bluestore: fix use after free race with aio\_wait ([pr#14956](#), Sage Weil)
- bluestore: os/bluestore: fix wal-queue bytes-counter to keep pace with others ([pr#13382](#), xie xingguo)
- bluestore: os/bluestore: fsck: verify blob.unused field ([pr#14316](#), Sage Weil)
- bluestore: os/bluestore: handle rounding error in cache ratios ([pr#15672](#), Sage Weil)
- bluestore: os/bluestore: implement collect\_metadata ([pr#14115](#), Sage Weil)
- bluestore: os/bluestore: include logical object offset in crc error ([pr#13074](#), Sage Weil)
- bluestore: os/bluestore: initialize finishers properly ([pr#15666](#), xie xingguo)
- bluestore: os/bluestore/KernelDevice: fix comments ([pr#15264](#), xie xingguo)
- bluestore: os/bluestore/KernelDevice: fix debug message ([pr#13135](#), Sage Weil)
- bluestore: os/bluestore/KernelDevice: helpful warning when aio limit exhausted ([pr#15116](#), Sage Weil)
- bluestore: os/bluestore/KernelDevice: kill zeros ([pr#12856](#), xie xingguo)
- bluestore: os/bluestore: kill BufferSpace.empty() ([pr#12871](#), xie xingguo)
- bluestore: os/bluestore: kill orphan declaration of do\_write\_check\_depth() ([pr#12853](#), xie xingguo)
- bluestore: os/bluestore: leverage the type knowledge in BitMapAreaLeaf ([pr#13736](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: Make BitmapFreelistManager kv iterator short lived ([pr#16243](#), Mark Nelson)
- bluestore: os/bluestore: make live changes for BlueStore throttle config work like initial config ([pr#14225](#), J. Eric Ivancich)

- bluestore: os/bluestore: miscellaneous fixes to BitAllocator ([pr#12696](#), xie xingguo)
- bluestore: os/bluestore: misc fix and cleanups ([pr#16315](#), Jianpeng Ma)
- bluestore: os/bluestore: misc fixes ([pr#14333](#), Sage Weil)
- bluestore: os/bluestore: move aio.h/cc from fs dir to bluestore dir ([pr#16409](#), Pan Liu)
- bluestore: os/bluestore: move object exist in assign nid ([pr#16117](#), Jianpeng Ma)
- bluestore: os/bluestore: move sharedblob to new collection in same shard ([issue#20358](#), [pr#15783](#), Sage Weil)
- bluestore: os/bluestore: narrow cache lock range; make sure min\_alloc\_size p2 aligned ([pr#15911](#), xie xingguo)
- bluestore: os/bluestore: “noid” is not always necessary in clone op ([pr#13769](#), wangzhengyong)
- bluestore: os/bluestore: nullptr in OmapIteratorImpl::valid ([pr#12900](#), Xinze Chi)
- bluestore: os/bluestore/NVMEDevice: Add multiple thread support for SPDK I/O thread ([pr#14420](#), Ziye Yang)
- bluestore: os/bluestore/NVMEDevice.cc: fix the random read issue ([pr#13055](#), optimistyzy)
- bluestore: os/bluestore/NVMEDevice: fix the compilation issue for collect\_metadata ([pr#14455](#), optimistyzy)
- bluestore: os/bluestore/NVMEdevice: fix the unrelease segs issue ([pr#12862](#), optimistyzy)
- bluestore: os/bluestore: only submit deferred if there is any ([pr#16269](#), Sage Weil)
- bluestore: os/bluestore: preallocate object[extent\_shard] key to avoid reallocate ([pr#12644](#), xie xingguo)
- bluestore: os/bluestore: pre-calculate number of ghost buffers to evict ([pr#15029](#), xie xingguo)
- bluestore: os/bluestore: put strings in mempool ([pr#12651](#), Allen Samuels, Sage Weil)
- bluestore: os/bluestore: Record l\_bluestore\_state\_kv\_queued\_lat for sync\_submit ([pr#14448](#), Jianpeng Ma)
- bluestore: os/bluestore: reduce some overhead for \_do\_clone\_range() and \_do\_remove() ([pr#15944](#), xie xingguo)

- bluestore: os/bluestore: refactor BlueStore::\_do\_write; kill dead ExtentMap::find\_lextent() method ([pr#15750](#), xie xingguo)
- bluestore: os/bluestore: refactor ExtentMap::update to avoid preceeding db updat... ([pr#12394](#), Igor Fedotov)
- bluestore: os/bluestore: remove a never read value ([pr#12618](#), liuchang0812)
- bluestore: os/bluestore: Remove ExtentFreeListManager ([pr#14772](#), Jianpeng Ma)
- bluestore: os/bluestore: remove intermediate key var to avoid string copy ([pr#12643](#), xie xingguo)
- bluestore: os/bluestore: remove no use parameter in bluestore\_blob\_t::map\_bt ([pr#13013](#), wangzhengyong)
- bluestore: os/bluestore: remove unneeded indirection in BitMapZone ([pr#13743](#), Radoslaw Zarzynski)
- bluestore: os/bluestore: remove unused condition variable ([pr#14973](#), Igor Fedotov)
- bluestore: os/bluestore: remove unused local variable "pos" ([pr#13715](#), wangzhengyong)
- bluestore: os/bluestore: remove unused variables ([pr#15718](#), zhanglei)
- bluestore: os/bluestore: rename/fix throttle options ([pr#14717](#), Sage Weil)
- bluestore: os/bluestore: roundoff bluefs allocs to bluefs\_alloc\_size ([pr#14876](#), Ramesh Chander)
- bluestore: os/bluestore: shrink buffer\_map key into uint32\_t ([pr#12850](#), xie xingguo)
- bluestore: os/bluestore: slightly refactor Blob::try\_reuse\_blob ([pr#15836](#), xie xingguo)
- bluestore: os/bluestore: some cleanup ([pr#13390](#), liuchang0812)
- bluestore: os/bluestore: space between func and contents ([pr#16804](#), xie xingguo)
- bluestore: os/bluestore: stop calculating bound if we must reshards; narrow shard combination condition ([pr#15631](#), xie xingguo)
- bluestore: os/bluestore/StupidAllocator: rounded down len to an align boundary ([issue#20660](#), [pr#16593](#), Zhu Shangzhong)
- bluestore: os/bluestore: target\_bytes should scale with meta/data ratios ([pr#15708](#), Mark Nelson)
- bluestore: os/bluestore: \_txc\_release\_alloc when do wal cleaning ([pr#12692](#), Xinze

Chi)

- bluestore: os/bluestore: use bufferlist functions whenever possible ([pr#16158](#), Jianpeng Ma)
- bluestore: os/bluestore: use correct bound encode size for unused ([pr#14731](#), Haomai Wang)
- bluestore: os/bluestore: use reference to avoid string copy ([pr#16364](#), Pan Liu)
- bluestore: os: extend ObjectStore interface to dump store's performance counters ([pr#13203](#), Igor Fedotov)
- bluestore,performance: common/config\_opts.h: compaction readahead for bluestore/rocksdb ([pr#14932](#), Mark Nelson)
- bluestore,performance: kv/RocksDBStore: implement rm\_range\_keys operator interface and test ([pr#13855](#), Haomai Wang)
- bluestore,performance: os/aio: remove the redundant memset(struct iocb) ([pr#13662](#), Jianpeng Ma)
- bluestore,performance: os/bluestore: add bluestore\_prefer\_wal\_size option ([pr#13217](#), Sage Weil)
- bluestore,performance: os/bluestore: avoid overloading extents during reshards; atomic deferred\_batch\_ops ([pr#15502](#), xie xingguo)
- bluestore,performance: os/bluestore: avoid the VTABLE-related burden in BitMapAllocator's hotspot ([pr#14348](#), Radoslaw Zarzynski)
- bluestore,performance: os/bluestore: batch throttle ([pr#15284](#), Jianpeng Ma)
- bluestore,performance: os/bluestore/BlueFS: add bluefs\_sync\_write option ([pr#14510](#), Sage Weil)
- bluestore,performance: os/bluestore/BlueFS: optimize get\_allocated ([pr#14121](#), Jianpeng Ma)
- bluestore,performance: os/bluestore/BlueFS: tune flushing of writes ([pr#13032](#), Sage Weil)
- bluestore,performance: os/bluestore/bluestore\_types: drop std::bitset for blob unused ([pr#12569](#), Sage Weil)
- bluestore,performance: os/bluestore: cap rocksdb cache size ([pr#15786](#), Mark Nelson)
- bluestore,performance: os/bluestore: default 16KB min\_alloc\_size on ssd ([pr#14076](#), Sage Weil)
- bluestore,performance: os/bluestore: default cache size of 3gb ([pr#15976](#), Sage Weil)

Weil)

- bluestore,performance: os/bluestore: differ default cache size for hdd/ssd backends ([pr#16157](#), xie xingguo)
- bluestore,performance: os/bluestore: do not balance bluefs on every kv\_sync\_thread iteration ([pr#14557](#), Sage Weil)
- bluestore,performance: os/bluestore: do not cache shard keys ([pr#12634](#), Sage Weil)
- bluestore,performance: os/bluestore: eliminate some excessive stuff ([pr#14675](#), Igor Fedotov)
- bluestore,performance: os/bluestore: fix deferred writes; improve flush ([pr#13888](#), Sage Weil)
- bluestore,performance: os/bluestore: generate same onode extent-shard keys in a more efficient way ([pr#12681](#), xie xingguo)
- bluestore,performance: os/bluestore: get rid off excessive lock at BitMapAllocator ([pr#14749](#), Igor Fedotov)
- bluestore,performance: os/blueStore: In osd\_tp\_thread, call \_txc\_finalize\_kv ([pr#14709](#), Jianpeng Ma)
- bluestore,performance: os/bluestore: keep statfs replica in RAM to avoid expensive KV retrieval ([pr#15309](#), Igor Fedotov)
- bluestore,performance: os/bluestore/KernelDevice: batch aio submit ([pr#16032](#), Haodong Tang)
- bluestore,performance: os/bluestore/KernelDevice: fix sync write vs flush ([pr#15034](#), Sage Weil)
- bluestore,performance: os/bluestore: kvdb histogram ([pr#12620](#), Varada Kari)
- bluestore,performance: os/bluestore: make bluestore\_max\_blob\_size parameter hdd/ssd case dependant ([pr#14434](#), Igor Fedotov)
- bluestore,performance: os/bluestore: memory and dereference clean-up in the BitAllocator ([pr#13811](#), Radoslaw Zarzynski)
- bluestore,performance: os/bluestore: move cache\_trim into MempoolThread ([pr#15380](#), xie xingguo)
- bluestore,performance: os/bluestore: optimize blob usage when doing appends/overwrites ([pr#13337](#), Igor Fedotov)
- bluestore,performance: os/bluestore: optimized (encode|decode)\_escaped ([pr#15759](#), Piotr Dałek)

- bluestore,performance: os/bluestore: partial reshards support ([pr#13162](#), Sage Weil)
- bluestore,performance: os/bluestore: prevent lock for almost “flush” calls ([pr#12524](#), Haomai Wang)
- bluestore,performance: os/bluestore: put bluefs in the middle of the shared device ([pr#14873](#), Sage Weil)
- bluestore,performance: os/bluestore: refactor small write handling to reuse blob more effect... ([pr#14399](#), Igor Fedotov)
- bluestore,performance: os/bluestore: remove CephContext\* from BmapEntry ([pr#13651](#), Radoslaw Zarzynski)
- bluestore,performance: os/bluestore: replace Blob ref\_map with reference counting ([pr#12904](#), Igor Fedotov)
- bluestore,performance: os/bluestore: rewrite deferred write handling ([issue#16644](#), [pr#14491](#), Sage Weil)
- bluestore,performance: os/bluestore: separate kv\_sync\_thread into two parts ([pr#14035](#), Jianpeng Ma, Igor Fedotov, Sage Weil)
- bluestore,performance: os/bluestore: set cache meta ratio to .9 ([pr#12635](#), Sage Weil)
- bluestore,performance: os/bluestore: the exhausted check in BitMapZone can be lock-less ([pr#13653](#), Radoslaw Zarzynski)
- bluestore,performance: os/bluestore: try to unshare blobs for EC overwrite workload ([pr#14239](#), Sage Weil)
- bluestore,performance: os/bluestore: tune deferred\_batch\_ops separately for hdd and ssd ([pr#14435](#), Sage Weil)
- bluestore,performance: os/bluestore: unify throttling model ([issue#19542](#), [pr#14306](#), Sage Weil)
- bluestore,performance: os/bluestore: use aio for reads ([issue#19030](#), [pr#13066](#), Sage Weil)
- bluestore,performance: os/bluestore: use Best-Effort policy when evicting onode from cache ([pr#12876](#), xie xingguo)
- bluestore,performance: os/bluestore: use denc for varint encoding ([pr#14911](#), Piotr Dałek)
- bluestore,performance: os/bluestore: various onode changes to reduce its in-memory footprint ([pr#12700](#), Igor Fedotov)

- bluestore, performance: os/fs/aio: use small\_vector for aio\_t; clean up header location ([pr#14853](#), Sage Weil)
- bluestore: rocksdb: add option: writable\_file\_max\_buffer\_size = 0 ([pr#12562](#), Jianpeng Ma)
- bluestore, tests: ceph-dencoder: enable bluestore types ([pr#13595](#), Willem Jan Withagen, Kefu Chai)
- bluestore, tests: ceph\_test\_objectstore: match clone\_range src and dst offset ([pr#13211](#), Sage Weil)
- bluestore, tests: qa/objectstore/bluestore\*: fsck on mount ([pr#15785](#), Sage Weil)
- bluestore, tests: test/ceph-test-objectstore: Don't always include BlueStore code ([pr#13516](#), Willem Jan Withagen)
- bluestore, tests: test/objectstore/store\_test\_fixture.cc: Exclude bluestore code if required ([pr#14085](#), Willem Jan Withagen)
- bluestore, tests: test/store\_test: add deferred test case setup to support explicit min... ([issue#18857](#), [pr#13415](#), Igor Fedotov)
- bluestore, tests: test/store\_test: fix bluestore test cases disablement ([pr#14228](#), Igor Fedotov)
- bluestore, tests: test/unittest\_bluefs: check whether add\_block\_device success ([pr#14013](#), shiqi)
- bluestore, tests: test/unittest\_bluefs: When fsync ret is less than 0, fsync can not be... ([pr#15365](#), shiqi)
- bluestore, tests: unittest\_alloc: add test\_alloc\_big ([issue#16662](#), [pr#14844](#), Sage Weil)
- bluestore, tools: ceph-bluestore-tool: rename from bluefs-tool; improve usage ([pr#14258](#), Sage Weil)
- bluestore, tools: ceph-kvstore-tool: allow 'bluestore-kv' as kvdb type; add escaping, compaction ([pr#14718](#), Sage Weil)
- bluestore: wrap blob id when it reaches maximum value of int16\_t ([issue#19555](#), [pr#15654](#), Xiaoyan Li)
- build/ops: 12.0.3 ([pr#15600](#), Jenkins Build Slave User)
- build/ops: add 12.0.1 release tag in master ([pr#14690](#), Jenkins Build Slave User)
- build/ops: add psmisc dependency to ceph-base (deb and rpm) ([issue#19129](#), [pr#13744](#), Nathan Cutler)
- build/ops: add sanity checks to run-make-check.sh ([pr#12683](#), Nathan Cutler)

- build/ops: alpine: add alpine linux dev support ([pr#9853](#), John Coyle)
- build/ops: arch: fix build on PowerPC with FreeBSD ([pr#14378](#), Andrew Solomon)
- build/ops: arch: fix cmake's ARM CRC intrinsics test to handle duplicitous gcc 4.8.5 ([issue#19386](#), [pr#14132](#), Dan Mick)
- build/ops: arch: use \_\_get\_cpuid instead of do\_cpuid ([issue#7869](#), [pr#14857](#), Jos Collin)
- build/ops: auth: Let's not use the deprecated cephx option ([pr#12721](#), Dave Chen)
- build/ops: build: Add Virtuozzo Linux support ([pr#14301](#), Andrey Parfenov)
- build/ops: build: build erasure-code isa lib without versions ([pr#16205](#), James Page)
- build/ops: build/cmake: provide asan, tsan, ubsan builds ([pr#12615](#), Matt Benjamin)
- build/ops: build: execute dh\_systemd\_{enable,start} after dh\_install ([issue#19585](#), [pr#16218](#), James Page)
- build/ops: build: move bash\_completion.d/ceph to ceph-common ([pr#15148](#), Leo Zhang)
- build/ops: build: remove ceph-disk-udev entirely ([pr#15259](#), Leo Zhang)
- build/ops: build: remove ceph-qa-suite directory ([pr#13880](#), Casey Bodley)
- build/ops: build: revert -Wvla from #15342 ([pr#15469](#), Willem Jan Withagen)
- build/ops: builds with dpdk v16.07 ([pr#12707](#), Kefu Chai)
- build/ops: build: Use .S suffix for ppc64le assembly files ([issue#20106](#), [pr#15373](#), Andrew Solomon)
- build/ops: ceph-disk: ability to use a different cluster name with dmcrypt ([issue#17821](#), [pr#11786](#), Sébastien Han, Erwan Velu)
- build/ops: ceph-disk: don't activate suppressed journal devices ([issue#19489](#), [pr#16123](#), David Disseldorp)
- build/ops: ceph.in: allow developer mode from outside build tree ([issue#20472](#), [pr#16055](#), Dan Mick)
- build/ops: ceph\_release: we are in the 'rc' phase (12.1.z) ([pr#15957](#), Sage Weil)
- build/ops: ceph.spec.in, debian/control: Add bc to build dependencies ([issue#18876](#), [pr#13338](#), Kyr Shatskyy)
- build/ops: Clean up make check for persistent test nodes (like arm64) ([pr#16773](#),

Dan Mick)

- build/ops: cmake,crc32c: conditionalize crc32c on different archs ([pr#14289](#), Kefu Chai)
- build/ops: CMakeLists.txt: boost\_python.so requires libpython.\*.so on FreeBSD ([pr#12763](#), Willem Jan Withagen)
- build/ops: CMakeLists.txt: don't do crypto/isa-l if not Intel ([pr#14721](#), Dan Mick)
- build/ops: CMakeLists.txt: suppress unneeded warning about jemalloc ([pr#13377](#), Willem Jan Withagen)
- build/ops,common: build: Adds C++ warning flag for C Variable-Length Arrays ([pr#15342](#), Jesse Williamson)
- build/ops,common: common/blkdev.cc: propagate get\_device\_by\_fd to different OSes ([pr#15547](#), Willem Jan Withagen)
- build/ops: common/module.c: do not use strerror\_r the GNU way ([pr#12363](#), Willem Jan Withagen)
- build/ops: compressor/zstd: add zstd to embedded ceph ([pr#13159](#), Bassam Tabbara)
- build/ops: conditionalize rgw Beast frontend so it isn't built on s390x architecture ([issue#20048](#), [pr#15225](#), Willem Jan Withagen, Nathan Cutler, Kefu Chai, Tim Serong, Casey Bodley)
- build/ops,core: build: let FreeBSD build ceph-fuse ([pr#14282](#), Willem Jan Withagen)
- build/ops,core: ceph-disk: use correct user in check\_journal\_req ([issue#18538](#), [pr#12947](#), Samuel Matzek)
- build/ops,core: common/freebsd\_errno.cc: fix missing ([pr#15741](#), Willem Jan Withagen)
- build/ops,core: erasure-code: update ec\_isa version + add missing AVX512 ISA-L sources ([pr#15636](#), Ganesh Mahalingam, Tushar Gohad)
- build/ops,core: os: allow offline conversion of filestore -> bluestore (or anything else) ([pr#14210](#), Sage Weil)
- build/ops,core: osd/OSD: auto class on osd start up ([pr#16014](#), xie xingguo)
- build/ops,core: osd/Pool: Disallow enabling 'hashpspool' option to a pool without '-yes-i-really-mean-it' ([issue#18468](#), [pr#13406](#), Vikhyat Umrao)
- build/ops,core,tests: osd/dmclock/testing: reorganize testing, building now optional ([issue#19987](#), [pr#15375](#), J. Eric Ivancich)

- build/ops: debian: Add missing tp files in deb packaging ([pr#13526](#), Ganesh Mahalingam)
- build/ops: debian: ceph-mgr: fix package description ([pr#15513](#), Fabian Grünbichler)
- build/ops: debian/control: add ceph-base-dbg ([pr#13796](#), Sage Weil)
- build/ops: debian: drop boost build dependencies ([pr#13524](#), Kefu Chai)
- build/ops: debian: package ceph.logroate properly ([issue#19390](#), [pr#14600](#), Kefu Chai)
- build/ops: debian: package crypto plugin only on amd64 ([pr#14820](#), Kefu Chai)
- build/ops: debian/rpm: move radosgw-admin to ceph-common ([issue#19577](#), [pr#14940](#), Ali Maredia)
- build/ops: debian/rules, ceph.spec.in: invoke cmake with -DBOOST\_J ([pr#14114](#), Dan Mick)
- build/ops: debian: sync logrotate packaging with downstream ([issue#19938](#), [pr#15567](#), Fabian Grünbichler)
- build/ops: debian: workaround the bug in dpkg-maintscript-helper ([issue#20453](#), [pr#16072](#), Kefu Chai)
- build/ops: debian: wrap-and-sort all files ([pr#16110](#), James Page)
- build/ops: dmclock: error: 'function' in namespace 'std' does not name a template type ([pr#14909](#), Jos Collin)
- build/ops: dmclock: include missing <functional> header ([pr#14923](#), Jos Collin)
- build/ops: dmclock: initial commit of dmclock QoS library ([pr#14330](#), J. Eric Ivancich)
- build/ops: do\_cmake.sh: enable ccache if installed ([pr#15274](#), Sage Weil)
- build/ops: do\_cmake.sh: fix syntax for /bin/sh (doesn't have +=) ([pr#16433](#), Dan Mick)
- build/ops: do\_freebsd.sh: Remove ENODATA requirement ([pr#13626](#), Willem Jan Withagen)
- build/ops: drop libfcgi build dependency ([pr#15285](#), Nathan Cutler)
- build/ops: gitignore: Ignore rejects by patch ([pr#14405](#), Willem Jan Withagen)
- build/ops: include/assert: test c++ before using static\_cast<> ([pr#16424](#), Kefu Chai)

- build/ops: init-ceph: add ceph libraries path to environment ([pr#14693](#), Mohamad Gebai)
- build/ops: init-ceph: fix ceph user args ([pr#13467](#), Sage Weil)
- build/ops: init-ceph: Make init-ceph work under FreeBSD for init-system ([pr#13209](#), Willem Jan Withagen)
- build/ops: install-deps.sh: add missing dependencies for FreeBSD ([pr#16545](#), Alan Somers)
- build/ops: install-deps.sh: workaround setuptools' dependency on six ([pr#15406](#), Kefu Chai)
- build/ops: mailmap: Update OVH contributors ([pr#13063](#), Bartłomiej Świecki)
- build/ops: make package groups comply with openSUSE guidelines ([issue#19184](#), [pr#13781](#), Nathan Cutler)
- build/ops: make-srpm: Pass first parameter to make-dist for building SRPM ([pr#13480](#), Wido den Hollander)
- build/ops: merge v12.0.2 release tag ([pr#15091](#), Jenkins Build Slave User)
- build/ops,mgr: debian/ceph-base.dirs: create bootstrap-mgr dirs ([pr#14838](#), Sage Weil)
- build/ops: miscellaneous cleanups and fixes (run-make-check.sh, ceph.spec.in) ([issue#20091](#), [issue#20127](#), [pr#15399](#), Nathan Cutler)
- build/ops,mon: mon/ConfigKeyService: add 'config-key dump' to show keys and vals ([pr#14858](#), Dan Mick)
- build/ops,mon: systemd: Restart Mon after 10s in case of failure ([issue#18635](#), [pr#13057](#), Wido den Hollander)
- build/ops: msg/async/rdma: compile with rdma as default ([pr#13901](#), DanielBar-On)
- build/ops: os/bluestore: fix build errors when spdk is on ([pr#16118](#), Ilsoo Byun)
- build/ops: packaging: install libceph-common.so\* not libceph-common.so.\* ([issue#18692](#), [pr#13148](#), Kefu Chai)
- build/ops,performance: crc32c: Add crc32c function optimized for ppc architecture ([pr#13909](#), Andrew Solomon)
- build/ops,performance,rbd: byteorder: use gcc intrinsics for byteswap ([pr#15012](#), Kefu Chai)
- build/ops,rbd,rgw: CMakeLists: trim rbd/rgw forced dependencies ([pr#16574](#), Patrick Donnelly)

- build/ops,rbd,tests: test/librbd: decouple ceph\_test\_librbd\_api from libceph-common ([issue#20175](#), [pr#15611](#), Kefu Chai)
- build/ops,rbd,tests: test/librbd: re-enable internal tests in ceph\_test\_librbd ([pr#16255](#), Mykola Golub)
- build/ops,rbd,tests: test: Need to exclude the fsx executable also on FreeBSD ([pr#13686](#), Willem Jan Withagen)
- build/ops: Revert “msg/async: increase worker reference with local listen table enabled backend” ([issue#20603](#), [pr#16323](#), Haomai Wang)
- build/ops: Revert “msg/async/rdma: Debug prints for ibv” ([pr#14245](#), Kefu Chai)
- build/ops,rgw: rgw\_file: radosgw-admin can be built under FreeBSD ([pr#12191](#), Willem Jan Withagen)
- build/ops,rgw,tests,tools: vstart: allow to start multiple radosgw when RGW=x ([pr#15632](#), Adam Kupczyk)
- build/ops,rgw,tools: vstart: add -rgw\_compression to set rgw compression plugin ([pr#15929](#), Casey Bodley)
- build/ops: rocksdb: build with ppc64 ([pr#12908](#), Kefu Chai)
- build/ops: rocksdb: sync with upstream ([pr#14456](#), Kefu Chai)
- build/ops: rocksdb: sync with upstream ([pr#14818](#), Nathan Cutler, Kefu Chai)
- build/ops: rpm: apply epoch only if %epoch macro is defined ([pr#15286](#), Nathan Cutler)
- build/ops: rpm: build ceph-resource-agents by default ([issue#17613](#), [pr#13515](#), Nathan Cutler)
- build/ops: rpm: bump epoch ahead of RHEL base ([issue#20508](#), [pr#16126](#), Ken Dreyer)
- build/ops: rpm,deb: fix ceph-volume ([issue#20915](#), [pr#16832](#), Sage Weil)
- build/ops: rpm: disable dwz to speed up valgrind ([issue#19099](#), [pr#13748](#), Kefu Chai)
- build/ops: rpm: drop boost build dependencies ([pr#13519](#), Nathan Cutler)
- build/ops: rpm: Drop legacy libxio support ([pr#16449](#), Nathan Cutler)
- build/ops: rpm: fix python-Sphinx package name for SUSE ([pr#15015](#), Nathan Cutler, Jan Matejek)
- build/ops: rpm: fix typo WTIH\_BABELTRACE ([pr#16366](#), Nathan Cutler)
- build/ops: rpm: Fix undefined FIRST\_ARG ([issue#20077](#), [pr#16208](#), Boris Ranto)

- build/ops: rpm: gperftools-devel >= 2.4 ([issue#13522](#), [pr#14870](#), Nathan Cutler)
- build/ops: rpm: make librbd1 %post scriptlet depend on coreutils ([issue#20052](#), [pr#15231](#), Giacomo Comes, Nathan Cutler)
- build/ops: rpm: move \_epoch\_prefix below Epoch definition ([pr#15417](#), Nathan Cutler)
- build/ops: rpm: move RDMA and python-prettytables build dependencies to distro-conditional section ([pr#15200](#), Nathan Cutler)
- build/ops: rpm: obsolete libcephfs1 ([pr#16074](#), Nathan Cutler)
- build/ops: rpm: package COPYING, move sample ceph.conf to ceph-common ([pr#15596](#), Nathan Cutler)
- build/ops: rpm: package crypto on x86\_64 only ([pr#14779](#), Nathan Cutler)
- build/ops: rpm: put mgr python build dependencies in make\_check bcond ([issue#20425](#), [pr#15940](#), Nathan Cutler, Tim Serong)
- build/ops: rpm: sane packaging of %{\_docdir}/ceph directory ([pr#15900](#), Nathan Cutler)
- build/ops: script: adding contributor credits script ([pr#13251](#), Patrick McGarry)
- build/ops: script: drop the -x arg for credits script ([pr#14296](#), Abhishek Lekshmanan)
- build/ops: script/sepio\_bt.sh: download packages from shaman not gitbuilder ([pr#12799](#), Kefu Chai)
- build/ops: script/sepio\_bt.sh: get sha1,release from t.log if it's not in core ([pr#13620](#), Kefu Chai)
- build/ops: script/sepio\_bt.sh: support xenial ([pr#13292](#), Kefu Chai)
- build/ops: selinux: Allow ceph daemons to read net stats ([issue#19254](#), [pr#13945](#), Boris Ranto)
- build/ops: selinux: Allow read on var\_run\_t ([issue#16674](#), [pr#15523](#), Boris Ranto)
- build/ops: selinux: Do parallel relabel on package install ([issue#20077](#), [pr#14871](#), Boris Ranto)
- build/ops: selinux: Install ceph-base before ceph-selinux ([issue#20184](#), [pr#15490](#), Boris Ranto)
- build/ops: Set subman cron attributes in spec file ([issue#20074](#), [pr#15270](#), Thomas Serlin)
- build/ops: spdk: upgrade spdk to v16.12 ([pr#12734](#), Pan Liu)

- build/ops: src/CMakeLists.txt: disable -Werror on rocksdb ([pr#12560](#), Willem Jan Withagen)
- build/ops: src/CMakeLists.txt: Move parse\_secret\_objs setting within definition block ([pr#12785](#), Willem Jan Withagen)
- build/ops: src/init-ceph.in: allow one((re)?start|stop) as commands ([pr#14560](#), Willem Jan Withagen)
- build/ops: sync luminous tag back to master ([pr#16758](#), Jenkins Build Slave User)
- build/ops: systemd: Add explicit Before=ceph.target ([pr#15835](#), Tim Serong)
- build/ops: systemd/ceph-disk: make it possible to customize timeout ([issue#18740](#), [pr#13197](#), Alexey Sheplyakov)
- build/ops: systemd/ceph-mgr: remove automagic mgr creation hack ([issue#19994](#), [pr#16023](#), Sage Weil)
- build/ops: systemd: remove ceph-create-keys from presets ([pr#14226](#), Sébastien Han)
- build/ops: systemd: Start OSDs after MONs ([issue#18516](#), [pr#13097](#), Boris Ranto)
- build/ops: test/fio\_ceph\_objectstore: fix fio plugin build failure caused by rec... ([pr#12655](#), Igor Fedotov)
- build/ops,tests: qa: make run-standalone work on FreeBSD ([pr#16595](#), Willem Jan Withagen)
- build/ops,tests: test/osd/CMakeLists.txt: osd-dup.sh require BlueStore/AIO ([pr#14387](#), Willem Jan Withagen)
- build/ops,tests: test/osd/osd-dup.sh: warn on low open file limit ([pr#14637](#), Piotr Dałek)
- build/ops,tests,tools: vstart.sh: Work around mgr restfull not available ([pr#15877](#), Willem Jan Withagen)
- build/ops: The Clangtastic Mr. Clocks ([pr#15186](#), Adam C. Emerson)
- build/ops: tool: add some ceph relate processes to ps-ceph.pl ([pr#12406](#), songbaisen)
- build/ops: tools/scripts:"FreeBSD getopt is not compatible, use the one from packages" ([pr#13260](#), Willem Jan Withagen)
- build/ops: tracing: Fix error in including all files in osd\_tp ([pr#12501](#), Ganesh Mahalingam)
- build/ops: upstart: start radosgw-all according to runlevel ([issue#18313](#), [pr#12586](#), Ken Dreyer)

- build/ops: vstart: clean up usage a bit ([pr#13138](#), Sage Weil)
- build/ops: vstart: do not start mgr if not start\_all ([pr#13974](#), Kefu Chai)
- build/ops: yasm-wrapper: filter -pthread ([pr#15249](#), Alessandro Barbieri)
- build/ops: yasm-wrapper: strip -E (stops ccache trashing source files) ([pr#14633](#), Tim Serong)
- cephfs: #11950: Persistent purge queue ([issue#11950](#), [pr#12786](#), John Spray)
- cephfs: #17980: MDS client blacklisting and blacklist on eviction ([issue#17980](#), [issue#9754](#), [pr#14610](#), John Spray)
- cephfs: #18600: Clear out tasks that don't make sense from multimds suite ([issue#18600](#), [pr#13089](#), John Spray)
- cephfs: ceph\_fuse: fix daemonization when pid file is non-empty ([pr#13532](#), "Yan, Zheng")
- cephfs: ceph\_fuse: pid\_file default to empty ([issue#18309](#), [pr#12628](#), Nathan Cutler)
- cephfs: ceph-fuse: use user space permission check by default ([issue#19820](#), [pr#14907](#), "Yan, Zheng")
- cephfs: ceph: simplify CInode::maybe\_export\_pin() ([pr#15106](#), "Yan, Zheng")
- cephfs: client: avoid returning negative space available ([issue#20178](#), [pr#15481](#), John Spray)
- cephfs: client: call the lru\_remove() twice, when trim cache ([pr#15662](#), huanwen ren)
- cephfs: client: check for luminous MDS before sending FLUSH\_MDLOG ([pr#15805](#), John Spray)
- cephfs: client/Client.cc: after reset session from MDS - reconnect ([issue#18757](#), [pr#13522](#), Henrik Korkuc)
- cephfs: client/Client.cc: prevent segfaulting ([issue#9935](#), [pr#12550](#), Michal Jarzabek)
- cephfs: client: client\_quota no longer optional ([pr#14978](#), Dan van der Ster)
- cephfs: client: don't request lookup parent if ino is root ([pr#12478](#), huanwen ren)
- cephfs: client: drop cap snaps when auth mds session gets closed ([issue#19022](#), [pr#13579](#), "Yan, Zheng")
- cephfs: client: fix clang warn of "argument is an uninitialized value" ([pr#12580](#),

- liuchang0812)
- cephfs: client: fix Client::handle\_cap\_flushsnap\_ack() crash ([issue#18460](#), [pr#12859](#), Yan, Zheng)
  - cephfs: client: fix Dentry::dump ([pr#15779](#), huanwen ren)
  - cephfs: client: fix display ino in the ldout ([pr#15314](#), huanwen ren)
  - cephfs: client: fix potential buffer overflow ([pr#12515](#), Yunchuan Wen)
  - cephfs: client: fix the cross-quota rename boundary check conditions ([pr#12489](#), Greg Farnum)
  - cephfs: client: fix UserPerm::gid\_in\_group() ([issue#19903](#), [pr#15039](#), "Yan, Zheng")
  - cephfs: client: getattr before returning quota/layout xattrs ([issue#17939](#), [pr#14018](#), John Spray)
  - cephfs: client/inode: fix the dump type of Inode::dump() ([pr#15198](#), huanwen ren)
  - cephfs: client: populate metadata during mount ([issue#18361](#), [pr#12915](#), John Spray)
  - cephfs: client: priority to verify the correctness of the "flag" ([pr#12897](#), huanwen ren)
  - cephfs: client: refine fsync/close writeback error handling ([pr#14589](#), John Spray)
  - cephfs: client: remove dead log code ([pr#13093](#), Patrick Donnelly)
  - cephfs: client: remove request from session->requests when handling forward ([issue#18675](#), [pr#13124](#), "Yan, Zheng")
  - cephfs: client: simplify remove\_cap interface ([pr#12161](#), John Spray)
  - cephfs: client: specify inode in get\_caps log message ([pr#13966](#), John Spray)
  - cephfs: client: wait for lastest osdmap when handling set file/dir layout ([issue#18914](#), [pr#13580](#), "Yan, Zheng")
  - cephfs,common: common/MemoryModel: Bump int to long and drop mallinfo ([pr#13453](#), Xiaoxi Chen)
  - cephfs,common,core: librados,osdc: kill ack vs commit distinction ([pr#12607](#), Sage Weil)
  - cephfs,common: include/fs\_types: fix unsigned integer overflow ([pr#12440](#), runsisi)

- cephfs,common,rbd: blkin: librbd trace hooks ([pr#15053](#), Victor Araujo, Jason Dillaman)
- cephfs,common,rbd: osdc: cache should ignore error bhs during trim ([issue#18436](#), [pr#12966](#), Jason Dillaman)
- cephfs,core: Add test for `is_hacky_ecoverwrites` in cephfs pool checks ([pr#13466](#), John Spray)
- cephfs,core: cleanup: use `std::make_shared` to replace new ([pr#12276](#), Yunchuan Wen)
- cephfs,core,mon: mon/MDSMonitor: fix segv when multiple MDSs raise same alert ([pr#16302](#), Sage Weil)
- cephfs: fix mount point break off problem after mds switch occurred ([issue#19437](#), [pr#14267](#), Guan yunfei)
- cephfs: fix `write_buf`'s `_len` overflow problem ([issue#19033](#), [pr#13587](#), Yang Honggang)
- cephfs: fs/ceph-fuse: normalize file open flags on the wire ([pr#14822](#), Jan Fajerski)
- cephfs: libcephfs.cc: fix memory leak ([pr#12557](#), Michal Jarzabek)
- cephfs: libcephfs: cleanups ([pr#12830](#), huanwen ren)
- cephfs: libcephfs: fix cct refcount constructing from rados ([pr#12831](#), John Spray)
- cephfs: mds/MDBalancer: remove useless `check_targets` and `hit_targets` logic from MDS balancer ([issue#20131](#), [pr#15407](#), Zhi Zhang)
- cephfs: mds/MDLog.cc Fix perf counter type for `jlat` ([pr#13449](#), Xiaoxi Chen)
- cephfs: mds/Server.cc: Don't evict a slow client if ([issue#17855](#), [pr#12935](#), Michal Jarzabek)
- cephfs: mds/StrayManager: avoid reusing deleted inode in `StrayManager::_purge_stray_logged` ([issue#18877](#), [pr#13347](#), Zhi Zhang)
- cephfs,mgr: pybind/mgr/fsstatus: use `mds_mem.dn` as dentry counter ([pr#15255](#), Zhi Zhang)
- cephfs: Mitigation for #16842, validate sessions after load ([issue#16842](#), [pr#14164](#), John Spray)
- cephfs: mon/FSCommand: fix indentation ([pr#15423](#), Sage Weil)
- cephfs: mon/MDSMonitor.cc: refuse fs new on pools with obj ([issue#11124](#), [pr#12825](#), Michal Jarzabek)

- cephfs: mon/MDSMonitor: respect mds\_standby\_for\_rank config ([pr#15129](#), “Yan, Zheng”)
- cephfs: mount: do not print “unknown” option to kclient ([issue#18159](#), [pr#12465](#), John Spray)
- cephfs: osdc/Filer: truncate large file party by party ([issue#19755](#), [pr#14769](#), “Yan, Zheng”)
- cephfs: osdc/Journaler: avoid executing on\_safe contexts prematurely ([issue#20055](#), [pr#15240](#), “Yan, Zheng”)
- cephfs: osdc/Journaler: fix memory leak in Journaler::issue\_read() ([issue#20338](#), [pr#15776](#), “Yan, Zheng”)
- cephfs: osdc/Objecter: fix inflight\_ops update ([pr#15768](#), “Yan, Zheng”)
- cephfs: osdc: remove journaler\_allow\_split\_entries option ([issue#19691](#), [pr#14636](#), John Spray)
- cephfs,performance: client: make seeky readdir more efficiency ([issue#19306](#), [pr#14317](#), “Yan, Zheng”)
- cephfs,performance: mds/server: skip unwanted dn in handle\_client\_readdir ([pr#12870](#), Xiaoxi Chen)
- cephfs: Permit recovering metadata into a new RADOS pool ([issue#15069](#), [issue#15068](#), [pr#10636](#), Douglas Fuller)
- cephfs: qa/cephfs: disable mds\_bal\_frag for TestStrays.test\_purge\_queue\_op\_rate ([issue#19892](#), [pr#15105](#), “Yan, Zheng”)
- cephfs: qa/cephfs: Fix for test\_data\_scan ([issue#19893](#), [pr#15094](#), Douglas Fuller)
- cephfs: qa: fix race in Mount.open\_background ([issue#18661](#), [pr#13137](#), John Spray)
- cephfs: qa/suites/fs: reserve more space for mds in full tests ([issue#19891](#), [pr#15026](#), “Yan, Zheng”)
- cephfs: qa/tasks/cephfs: use getattr to guarantee inode is in client cache ([issue#19912](#), [pr#15062](#), “Yan, Zheng”)
- cephfs: qa: unpin knfs from ubuntu ([issue#16397](#), [pr#13088](#), John Spray)
- cephfs: qa: update log whitelists for kcephfs suite ([pr#14922](#), “Yan, Zheng”)
- cephfs: qa: update remaining ceph.com to download.ceph.com ([issue#18574](#), [pr#12964](#), John Spray)
- cephfs: qa: whitelist new fullness messages in fs tests ([issue#19253](#), [pr#13915](#), John Spray)

- cephfs: Remove “experimental” warnings from multimds ([pr#15154](#), John Spray, “Yan, Zheng”)
- cephfs: Rewrite mount.fuse.ceph (to python) and move ceph-fuse options to fs\_mntops ([pr#11448](#), Edgaras Lukosevicius)
- cephfs: tasks/cephfs: fix kernel force umount ([issue#18396](#), [pr#12833](#), Yan, Zheng)
- cephfs: test/libcephfs: avoid buffer overflow when testing ceph\_getdents() ([issue#18941](#), [pr#13429](#), “Yan, Zheng”)
- cephfs,tests: Add multimds:thrash sub-suite and fix bugs in thrasher for multimds ([issue#18690](#), [issue#10792](#), [pr#13262](#), Patrick Donnelly)
- cephfs,tests: ceph-object-corpus: mark MMDSSlaveRequest incompat change ([pr#15730](#), Sage Weil)
- cephfs,tests: Improve vstart\_runner to (optionally) create its own cluster ([pr#12800](#), John Spray)
- cephfs,tests: qa: fix float parse error in test\_fragment ([pr#15122](#), Patrick Donnelly)
- cephfs,tests: qa: fix test\_standby\_for\_invalid\_fscid with vstart\_runner ([pr#14272](#), John Spray)
- cephfs,tests: qa: handle SSHException in logrotate ([pr#13359](#), John Spray)
- cephfs,tests: qa, mds: add checks for fragmentation, and enable it by default ([issue#16523](#), [pr#13862](#), john Spray, John Spray)
- cephfs,tests: qa: misc cephfs test improvements ([issue#20131](#), [pr#15411](#), John Spray)
- cephfs,tests: qa: re-enable ENOSPC tests for kclient ([issue#19550](#), [pr#14396](#), John Spray)
- cephfs,tests: qa: silence spurious insufficient standby health warnings ([pr#15035](#), Patrick Donnelly)
- cephfs,tests: qa: silence upgrade test failure ([issue#19934](#), [pr#15126](#), Patrick Donnelly)
- cephfs,tests: qa: simplify TestJournalRepair ([pr#15096](#), John Spray)
- cephfs,tests: qa/tasks: force umount during kclient teardown ([issue#18663](#), [pr#13099](#), John Spray)
- cephfs,tests: qa: Tidy up fs/ suite ([pr#14575](#), John Spray)
- cephfs,tests: qa/vstart\_runner: amend ps invocation ([pr#14254](#), Ilya Dryomov)

- cephfs, tests: qa: whitelist another fullness log message ([issue#19253](#), [pr#14221](#), John Spray)
- cephfs, tests: tasks/cephfs: tear down on mount() failure ([pr#13282](#), John Spray)
- cephfs: tools/cephfs: remove apply mode of cephfs-journal-tool ([pr#15715](#), John Spray)
- cephfs: tools/cephfs: set dir\_layout when injecting inodes ([issue#19406](#), [pr#14234](#), John Spray)
- ceph-volume: use unique logical volumes ([pr#17208](#), Alfredo Deza)
- cleanup: .gitignore: exclude rpm files ([pr#15745](#), Leo Zhang)
- cleanup: Move code from .h into .cc ([pr#12737](#), Amir Vadai)
- cleanup: resolve compiler warnings ([pr#13236](#), Adam C. Emerson)
- cleanup: src: put-to operator function - const input cleanup ([issue#3977](#), [pr#15364](#), Jos Collin)
- cmake: add “container” to required boost components ([pr#14850](#), Kefu Chai)
- cmake: Add -finstrument-functions flag to OSD code ([pr#15055](#), Mohamad Gebai)
- cmake: add RGW and MDS to libcephd ([pr#12345](#), Bassam Tabbara)
- cmake: Add simple recursive ctags target for Ceph source only ([pr#14334](#), Kefu Chai, Dan Mick)
- cmake: align cmake names of library packages ([issue#19853](#), [pr#14951](#), Nathan Cutler)
- cmake: Allow tests to build without NSS ([pr#13315](#), Daniel Gryniewicz)
- cmake: build boost as an external project ([pr#15376](#), Kefu Chai)
- cmake: build tracepoint libraries for vstart target ([pr#14354](#), Mohamad Gebai)
- cmake: check the existence of gperf before using it ([pr#15164](#), Kefu Chai)
- cmake: cleanup the use of udev and blkid in target\_link\_lib() ([pr#12811](#), Willem Jan Withagen)
- cmake: disable -fvar-tracking-assignments for config.cc ([pr#16695](#), Kefu Chai)
- cmake: disable mallinfo for jemalloc ([pr#12469](#), Bassam Tabbara)
- cmake: do not add dependencies to INTERFACE library on cmake < 3.3 ([pr#15813](#), Kefu Chai)
- cmake: do not compile crush twice ([pr#14725](#), Kefu Chai)

- cmake: do not link libcommon against some libs ([pr#15340](#), Willem Jan Withagen)
- cmake: do not try to add submodule to exclude list if .git is not around ([pr#14495](#), Kefu Chai)
- cmake: enable cross-compilation of boost ([issue#18938](#), [pr#14881](#), Kefu Chai)
- cmake: exclude \*.css while generating ctags ([pr#15663](#), Leo Zhang)
- cmake: explicitly call find\_package(PythonInterp) first to fix build err ([pr#12385](#), Yixun Lan)
- cmake: fix boost components for WITH\_SYSTEM\_BOOST ([pr#15160](#), Bassam Tabbara)
- cmake: Fix broken async/rdma compilation since move to libceph-common ([pr#13122](#), Oren Duer)
- cmake: fix broken RDMA compilation after merge PR #12878 ([pr#13186](#), Oren Duer)
- cmake: fix hard coded boost python lib ([pr#12480](#), John Coyle)
- cmake: fix rpath on shared libraries and binaries targets ([pr#12927](#), Ricardo Dias)
- cmake: fix the build with -DWITH\_ZFS=ON ([pr#15907](#), Kefu Chai)
- cmake: fix the linked lib reference of unittest\_rgw\_crypto ([pr#14869](#), Willem Jan Withagen)
- cmake: improved build speed by 5x when using ccache ([pr#15147](#), Bassam Tabbara)
- cmake: kill duplicated cmake commands ([pr#14948](#), liuchang0812)
- cmake: link against fcgi only if enabled ([pr#15425](#), Yao Zongyou)
- cmake: link ceph-{mgr,mon,mds,osd} against libcommon statically ([pr#12878](#), Kefu Chai)
- cmake: link consumers of libclient with libcommon ([issue#18838](#), [pr#13394](#), Kefu Chai)
- cmake: misc fixes for build on i386 ([pr#15516](#), James Page)
- cmake: pass -d0 to b2 if not CMAKE\_VERBOSE\_MAKEFILE ([pr#14651](#), Kefu Chai)
- cmake: remove Findpciaccess.cmake ([pr#12776](#), optimistyzy)
- cmake: Rewrite HAVE\_BABELTRACE option to WITH ([pr#15305](#), Willem Jan Withagen)
- cmake: rgw: do not link against boost in a wholesale ([pr#15347](#), Nathan Cutler, Kefu Chai)
- cmake: search for Keyutils in default paths ([pr#12769](#), Pascal Bach)

- cmake: search for nspr include files for both suffixes: nspr4 and nspr ([issue#18535](#), [pr#12939](#), John Lin)
- cmake: should not compile crc32c\_ppc.c on intel arch ([pr#14423](#), Kefu Chai)
- cmake: simplify find\_package jemalloc ([pr#12468](#), Bassam Tabbara)
- cmake: support for external rocksdb ([pr#12467](#), Bassam Tabbara)
- cmake: support optional argument for overriding default ctag excludes ([pr#14379](#), Kefu Chai)
- cmake: turn libcommon into a shared library ([pr#12840](#), Kefu Chai)
- cmake: use CMAKE\_INSTALL\_INCLUDEDIR ([pr#16483](#), David Disseldorp)
- cmake: workaround ccache issue with .S assembly files ([pr#15142](#), Bassam Tabbara)
- common: add ceph::size() ([pr#15181](#), Kefu Chai)
- common: add override in common and misc ([issue#18922](#), [pr#13443](#), liuchang0812)
- common: add override in header file ([pr#13774](#), liuchang0812)
- common: add override in msg subsystem ([pr#13771](#), liuchang0812)
- common: auth: Enhancement for the supported auth methods ([pr#12937](#), Dave Chen)
- common: auth/RotatingKeyRing: use std::move() to set secrets ([pr#15866](#), Kefu Chai)
- common: avoid statically allocating configuration options ([issue#20869](#), [pr#16735](#), Jason Dillaman)
- common: Better handling for missing/inaccessible ceph.conf files ([issue#19658](#), [pr#14757](#), Dan Mick)
- common: bufferlist: cleanup semantical wrong for bufferlist::append ([pr#12247](#), Yankun Li)
- common: buffer: silence unused var warning on FreeBSD ([pr#16452](#), Willem Jan Withagen)
- common: ceph\_osd: remove client message cap limit ([pr#14944](#), Haomai Wang)
- common: ceph: wait for maps before doing 'ceph tell ... help' ([issue#20113](#), [pr#16756](#), Sage Weil)
- common: cls/log/cls\_log.cc: reduce logging noise ([issue#19835](#), [pr#14879](#), Willem Jan Withagen)
- common: cls: optimize header file dependency ([pr#15165](#), Brad Hubbard, Xiaowei

Chen)

- common: cmdparse: more constness ([pr#15023](#), Kefu Chai)
- common: common/admin\_socket: add config for admin socket permission bits ([pr#11684](#), runsisi)
- common: common/admin-socket: fix potential buffer overflow ([pr#12518](#), Yunchuan Wen)
- common: common/auth: add override in headers ([pr#13692](#), liuchang0812)
- common: common/BackTrace: add operator<< ([pr#9028](#), Kefu Chai)
- common: common/BackTrace: demangle on FreeBSD also ([pr#12992](#), Kefu Chai)
- common: common/buffer: close pipe fd if set nonblocking fails ([pr#12828](#), donglinpeng)
- common: common/buffer: off-by-one error in max iov length blocking ([issue#20907](#), [pr#16803](#), Dan Mick)
- common: common/ceph\_context.cc: Use CEPH\_DEV to reduce logfile noise ([pr#10384](#), Willem Jan Withagen)
- common: common/ceph\_context: ‘config diff get’ option added ([pr#10736](#), Daniel Oliveira)
- common: common/ceph\_context: fewer warnings about experimental features ([pr#14170](#), Sage Weil)
- common: common/ceph\_context: fix leak of registered commands on exit ([pr#15302](#), xie xingguo)
- common: common/ceph\_context: Show clear message if all features are enabled ([pr#12676](#), Dave Chen)
- common: common/cmdparse.cc: remove unused variable ‘argnum’ in dump\_cmd\_to\_json() ([pr#16862](#), Luo Kexue)
- common: common/common\_init: disable default dout logging for UTILITY\_NODOUT too ([issue#20771](#), [pr#16578](#), Sage Weil)
- common: common/config: Add /usr/local/etc/ceph to default paths ([pr#14797](#), Willem Jan Withagen)
- common: common/config: eliminate config\_t::set\_val unsafe option ([issue#19106](#), [pr#13687](#), liuchang0812)
- common: common/config: fix return type of string::find and use string::npos ([pr#9924](#), Yan Jun)

- common: common,config: OPT\_FLOAT and OPT\_DOUBLE output format in config show ([issue#20104](#), [pr#15647](#), Yanhu Cao)
- common: common/config\_opt: remove unused config ([pr#15874](#), alex.wu)
- common: common/config\_opts: drop unused opt ([pr#15876](#), Yanhu Cao)
- common: common/config\_opts.h: FreeBSD timing changed due to no SO\_REUSEADDR ([pr#12594](#), Willem Jan Withagen)
- common: common/config\_opts.h: Remove deprecated osd\_compact\_leveldb\_on\_mount option ([issue#19318](#), [pr#14059](#), Vikhyat Umrao)
- common: common/config\_opts.h: remove obsolete configuration option ([pr#12659](#), Li Wang)
- common: common/config\_opts: Set the HDD throttle cost to 1.5M ([pr#14808](#), Mark Nelson)
- common: common/EventTrace: fix compiler warning ([pr#13659](#), Jianpeng Ma)
- common: common/Finisher: fix uninitialized variable warning ([pr#14958](#), Piotr Dałek)
- common: common/freebsd\_errno.cc: fixed again a stupid typo ([pr#15742](#), Willem Jan Withagen)
- common: common/interval\_set: return int64\_t for size() ([pr#12898](#), Xinze Chi)
- common: common/iso\_8601.cc: Make return expression Clang compatible ([pr#15336](#), Willem Jan Withagen)
- common: common/LogEntry: include EntityName in log entries ([pr#15395](#), Sage Weil)
- common: common/Mutex.cc: fixed the error in comment ([pr#16214](#), Pan Liu)
- common: common/options: refactors to set the properties in a more structured way ([pr#16482](#), Kefu Chai)
- common: common,osdc: remove atomic\_t completely ([pr#15562](#), Kefu Chai)
- common: common/perf\_counters: add average time for PERFCOUNTER\_TIME ([pr#15478](#), xie xingguo)
- common: common/perf\_counters: fix race condition with atomic variables ([pr#14227](#), J. Eric Ivancich)
- common: common/perf\_counters: make schema more friendly and update docs ([pr#14933](#), Sage Weil)
- common: common/perf\_counters.: Remove unnecessary judgment ([pr#10407](#), zhang.zezhu)

- common: common/simple\_spin: use \_\_ppc\_yield() on all powerpc archs ([pr#14310](#), Kefu Chai)
- common: common,test: migrate atomic\_t to std::atomic ([pr#14866](#), Jesse Williamson)
- common: common/Timer: do not add event if already shutdown ([issue#20432](#), [pr#16201](#), Kefu Chai)
- common: common/WorkQueue: use threadpoolname + threadaddr for heartbeat\_han... ([pr#16563](#), huangjun)
- common: common/xmlformatter: turn on underscored and add unittest ([pr#12916](#), liuchang0812)
- common: compressor/zlib: remove g\_ceph\_context/g\_conf from compressor plugin ([pr#16245](#), Casey Bodley)
- common: compressor/zstd: add zstd compression plugin ([pr#13075](#), Kefu Chai, Sage Weil)
- common: config: Improve warning for unobserved value ([issue#18424](#), [pr#12855](#), Brad Hubbard)
- common: config\_opt: use bool instead of int for the default value of filestore\_debug\_omap\_check ([pr#15651](#), Leo Zhang)
- common,core: ceph\_test\_rados\_api\_misc: fix LibRadosMiscConnectFailure.ConnectFailure retry ([issue#19901](#), [pr#15522](#), Sage Weil)
- common: core/common: Fix ENODATA for FreeBSD with compat.h ([issue#19883](#), [pr#15685](#), Willem Jan Withagen)
- common,core: common, osd, tools: Add histograms to performance counters ([pr#12829](#), Bartłomiej Święcki)
- common,core: common/pick\_address.cc: Copy public\_netw to cluster\_netw if cluster empty ([pr#12929](#), Willem Jan Withagen)
- common,core: common/TracepointProvider: add assert if dlopen error ([pr#13430](#), Jianpeng Ma)
- common,core: common/TrackedOp: make TrackedOp::reset\_desc() safe ([issue#19110](#), [pr#13702](#), Sage Weil)
- common,core: mempool: put bloom\_filter in mempool ([pr#13009](#), Sage Weil)
- common,core: osd,mds,mgr: do not dereference null rotating\_keys ([issue#20667](#), [pr#16455](#), Sage Weil)
- common,core: osd,osdc: pg and osd-based backoff ([pr#12342](#), Sage Weil)

- common,core: osd/OSDMap: make osd\_state 32 bits wide ([pr#15390](#), Sage Weil)
- common,core: osd/OSDMap: replace require\_osds flags with a single require\_osd\_release field ([pr#15068](#), Sage Weil)
- common,core: osd/OSDMap: replace string-based min\_compat\_client with a CEPH\_RELEASE uint8\_t ([pr#15351](#), Sage Weil)
- common,core: osd/osd\_types: add flag name (IGNORE\_REDIRECT) ([pr#15795](#), Myoungwon Oh)
- common,core: rados: we need to get the latest osdmap when pool does not exists ([pr#13289](#), song baisen)
- common,core,tests: Wip cppcheck errors ([pr#14446](#), Brad Hubbard)
- common: crc32c: include acconfig.h to fix ceph\_crc32c\_aarch64() ([pr#15515](#), Kefu Chai)
- common: crush/CrushWrapper: fix has\_incompat\_choose\_args ([pr#15218](#), Sage Weil)
- common: crush/CrushWrapper: fix has\_incompat\_choose\_args() ([pr#15244](#), Sage Weil)
- common: crypto: cleanup NSPR in main thread ([pr#14801](#), Kefu Chai)
- common: delete unused conf "filestore\_debug\_disable\_sharded\_check" ([pr#13051](#), Chuanhong Wang)
- common: denc: add encode/decode for basic\_sstring ([pr#15135](#), Kefu Chai, Casey Bodley)
- common: do not print error when asok is closed ([pr#14022](#), Patrick Donnelly)
- common: fix building against libcryptopp ([pr#14949](#), Shengjing Zhu)
- common: Fix clang compilation ([pr#13335](#), Bartłomiej Święcki)
- common: Fix heap buffer overflow in do\_request ([issue#19393](#), [pr#14173](#), Brad Hubbard)
- common: fix lockdep vs recursive mutexes ([pr#9940](#), Adam Kupczyk)
- common: fix log warnings ([pr#16056](#), xie xingguo)
- common: fix Option set\_long\_description ([pr#16668](#), Yan Jun)
- common: fix segfault in public IPv6 addr picking ([issue#19371](#), [pr#14124](#), Fabian Grünbichler)
- common: fix that \$host always expands to localhost instead of actual hostname ([issue#11081](#), [pr#12998](#), liuchang0812)

- common: fix typo in option of rados\_mon\_op\_timeout's comment ([pr#15681](#), Leo Zhang)
- common: Fix unused variable references warnings ([pr#14790](#), Willem Jan Withagen)
- common: follow up to new options infrastructure ([pr#16527](#), John Spray)
  - common: Forward-declare container I/O overloads
- common: get\_process\_name: use getprogname on bsd systems ([pr#15338](#), Mykola Golub)
- common: get rid of "warning: ignoring return value of 'strerror\_r'" ([pr#12775](#), xie xingguo)
- common: global: we need to handle the init\_on\_startup return value when global\_init ([pr#13018](#), song baisen)
- common: Implements simple\_spin\_t in terms of std::atomic\_flag ([pr#14370](#), Jesse Williamson)
- common: Improved CRC calculation for zero buffers ([pr#11966](#), Adam Kupczyk)
- common: include/ceph\_features.h uses uint64\_t, which is in sys/types.h ([pr#13339](#), Willem Jan Withagen)
- common: include/denc: improvements ([pr#12626](#), Adam C. Emerson)
- common: include/denc, kv: silence gcc warnings ([pr#13458](#), Kefu Chai)
- common: include/denc: remove nullptr runtime magic boundedness check ([pr#13889](#), Sage Weil)
- common: include/lru.h: add const to member functions ([pr#15408](#), yonghengdexin735)
- common: include/mempool: fix typo in comments ([pr#12772](#), huangjun)
- common: include/rados: Fix typo in rados\_ioctlx\_cct() doc ([pr#15220](#), Jos Collin)
- common: include: Redo some includes for FreeBSD ([issue#19883](#), [pr#15337](#), Willem Jan Withagen)
- common: initialize array in struct BackTrace ([pr#15864](#), Jos Collin)
- common: initialize \_hash in LogEntryKey() ([pr#15615](#), Jos Collin)
- common: int\_types.h: remove hacks to workaround old systems ([pr#15069](#), Kefu Chai)
- common: kv: resolve a crash issue in ~LevelDBStore() ([pr#16553](#), wumingqiao)
- common: librados, libradosstriper, test: migrate atomic\_t to std::atomic (baragon) ([pr#14658](#), Jesse Williamson)
- common: librados, osd: clang fixes ([pr#13768](#), Kefu Chai)

- common: libradosstriper: Add example code ([pr#15350](#), Logan Blyth)
- common: libradosstriper: fix format injection vulnerability ([issue#20240](#), [pr#15674](#), Stan K)
- common: libradosstriper: fix MultiAioCompletion leaks on failure ([pr#15471](#), Kefu Chai)
- common: make attempts of auth rotating configurable ([pr#12563](#), xie xingguo)
- common: Make spinlock delay more conventional ([pr#14248](#), Brad Hubbard)
- common: mempool: improve dump; fix buffer accounting bugs ([pr#15403](#), Sage Weil)
- common: messages: fix return type name of MOSDMap ([pr#14382](#), Leo Zhang)
- common: mgr/PyFormatter: implement dump\_format\_va ([pr#15634](#), Sage Weil)
- common: misc cleanups in common, global, os, osd submodules ([pr#16321](#), Yan Jun)
- common: misc fixes detected by crypto shutdown assert ([pr#12925](#), Sage Weil)
- common,mon: crush,mon: add weight-set introspection and manipulation commands ([pr#16326](#), Sage Weil)
- common,mon: messenger,client,compressor: migrate atomic\_t to std::atomic ([pr#14657](#), Jesse Williamson)
- common: mon/MonClient: scale backoff interval down when we have a healthy mon session ([issue#20371](#), [pr#16576](#), Kefu Chai, Sage Weil)
- common,mon: mon,crush: add 'osd crush swap-bucket' command ([pr#15072](#), Sage Weil)
- common: msg/async: add assert of ms\_async\_op\_threads > 0 ([pr#15629](#), linbing)
- common: msg/async: assert if compiled code doesn't support the configured ms ([pr#12559](#), Avner BenHanoch)
- common: msg/async: fix crash that writing char to nonblock-fd gets EAGAIN in EventCenter::wakeup ([pr#13822](#), liuchang0812)
- common: msg/async: make recv\_stamp more precise ([pr#15810](#), Pan Liu)
- common: msg/async/rdma: Add fork safe on RDMA ([pr#13740](#), Sarit Zubakov)
- common: msg/async/rdma: clean line endings ([pr#12688](#), Adir Lev)
- common: msg/async/rdma: Remove compilation warning ([pr#13142](#), Sarit Zubakov)
- common: msg/async/rdma: rename chunk\_size to buffer\_size ([pr#13666](#), Adir Lev)
- common: msg/async/rdma: Support for RoCE v2 and SL ([pr#12556](#), Oren Duer)

- common: msg/async/rdma: Update fix broken compilation ([pr#13940](#), Sarit Zubakov)
- common: msg/async: return right away in NetHandler::set\_priority() if not supported ([pr#14795](#), Kefu Chai)
- common: msg/simple: call clear\_pipe in wait() shutdown path ([issue#15784](#), [pr#12633](#), Sage Weil)
- common: msg/SimpleMessenger: error out misplace in set\_socket\_options ([pr#13961](#), wangzhengyong)
- common: msg/simple/Pipe: support IPv6 QoS ([issue#18887](#), [pr#13370](#), Robin H. Johnson)
- common: .organizationmap: Updated authors ([pr#14360](#), Jos Collin)
- common: osdc: fix osdc\_osd\_seesion perf counter ([pr#13478](#), Xiaoxi Chen)
- common: osdc/Objecter: fix bugs in explicit naming of op spg\_t ([pr#13534](#), Sage Weil)
- common: osdc/Objecter: fix pool dne corner case ([issue#19552](#), [pr#14901](#), Sage Weil)
- common: osdc/Objecter: handle command target that goes down ([issue#19452](#), [pr#14302](#), Sage Weil)
- common: osdc/Objecter: release message if it's not handled ([issue#19741](#), [pr#15890](#), Kefu Chai)
- common: osdc/Objecter: resend RWORDERED ops on full ([issue#19133](#), [pr#13759](#), Sage Weil)
- common: osd/OSDMap: fix feature commit comment ([pr#15056](#), Sage Weil)
- common: osd/OSDMap: get\_previous\_up\_osd\_before() may run into endless loop ([pr#12976](#), Mingxin Liu)
- common: osd/OSDMap: print require\_osd\_release ([pr#15974](#), Sage Weil)
- common: osd/osd\_types: clean up OSDOp printers ([pr#12980](#), Sage Weil)
- common: Passing null pointer option\_name to operator << in md\_config\_t::parse\_option() ([pr#15881](#), Jos Collin)
- common,performance: buffer: allow buffers to be accounted in arbitrary mempools ([pr#15352](#), Sage Weil)
- common,performance: common/Finisher: batch handle perfcounter && only send signal when waiter existed ([pr#14363](#), Jianpeng Ma)
- common,performance: crc32c: Add ppc64le fast zero optimized assembly ([pr#15100](#),

Andrew Solomon)

- common, performance: inline\_memory: optimized mem\_is\_zero for non-x64 ([pr#15307](#), Piotr Dałek)
- common, performance: kv/rocksdb: supports SliceParts interface ([pr#15058](#), Haomai Wang)
- common, performance: osd/OSDMap: make pg\_temp more efficient ([pr#15291](#), Sage Weil)
- common: possible lockdep false alarm for ThreadPool lock ([issue#18819](#), [pr#13258](#), Mykola Golub)
- common: prevent unset\_dumpable from generating warnings ([pr#16462](#), Willem Jan Withagen)
- common: rados: allow “rados purge” to delete objects when osd is full ([pr#13814](#), Pan Liu)
- common: rados: more info added to pool deletion error ([issue#19400](#), [pr#14235](#), Vedant Nanda)
- common, rbd: osdc/Objecter: unify disparate EAGAIN handling paths into one ([pr#16627](#), Sage Weil)
- common, rbd, rgw: common/escape: do not escape / in json ([pr#14130](#), Sage Weil)
- common, rbd, rgw: common/rgw/rbd: remove some unused variables ([pr#16690](#), Luo Kexue)
- common, rdma: msg/async/rdma: automatically set RDMAV\_HUGEPAGES\_SAFE according to conf ([pr#15755](#), DanielBar-On)
- common, rdma: msg/async/rdma: check ulimit ([pr#13655](#), Sarit Zubakov, Adir Lev)
- common, rdma: msg/async/rdma: Introduce Device.{cc,h} ([pr#14001](#), Amir Vadai)
- common, rdma: msg/async/rdma: Introduce RDMAConnMgr + Debug prints ([pr#14201](#), Amir Vadai)
- common, rdma: msg/async/rdma: Move resource handling to Device ([pr#14088](#), Sarit Zubakov, Amir Vadai)
- common, rdma: msg/async/rdma: RDMA-CM Initialize device on first connect ([pr#14179](#), Amir Vadai)
- common, rdma: msg/async/rdma: reduce number of rdma rx/tx buffers ([pr#13190](#), Adir Lev)
- common, rdma: msg/async/rdma: use lists properly ([pr#15908](#), Adir lev, Adir Lev)
- common: remove config opt conversion utility ([pr#16480](#), John Spray)

- common: remove n on clog messages ([pr#13794](#), Sage Weil)
- common: Remove redundant includes - 2 ([issue#19883](#), [pr#15169](#), Jos Collin)
- common: Remove redundant includes - 3 ([issue#19883](#), [pr#15204](#), Jos Collin)
- common: Remove redundant includes - 4 ([issue#19883](#), [pr#15251](#), Jos Collin)
- common: Remove redundant includes - 5 ([issue#19883](#), [pr#15267](#), Jos Collin)
- common: Remove redundant includes - 6 ([issue#19883](#), [pr#15299](#), Jos Collin)
- common: Remove redundant includes ([issue#19883](#), [pr#15003](#), Brad Hubbard)
- common: Remove redundant includes ([issue#19883](#), [pr#15019](#), Brad Hubbard)
- common: Remove redundant includes ([issue#19883](#), [pr#15042](#), Brad Hubbard)
- common: Remove redundant includes ([issue#19883](#), [pr#15086](#), Jos Collin)

- common: remove useless parameter ([pr#14096](#), baiyanchun)
- common: Revamp config option definitions ([issue#20627](#), [pr#16211](#), John Spray, Kefu Chai, Sage Weil)
- common, rgw: cls/refcount: store and use list of retired tags ([issue#20107](#), [pr#15673](#), Yehuda Sadeh)
- common: src/common/ceph\_string: stringify new osd states ([pr#15751](#), xie xingguo)
- common: src/common: change last\_work\_queue to next\_work\_queue ([pr#14738](#), Pan Liu)
- common: support s390 and unknown architectures in spin-wait loop ([issue#19492](#), [pr#14337](#), Nathan Cutler)
- common, tests: ceph\_test\_rados\_api\_c\_read\_operations: do not assert per-op rval is correct ([issue#19518](#), [pr#16196](#), Sage Weil)
- common, tests: ceph\_test\_rados\_api\_list: more fix LibRadosListNP.ListObjectsError ([issue#19963](#), [pr#15138](#), Sage Weil)
- common, tests: test: Make screencandy optional for FreeBSD ([pr#15444](#), Willem Jan Withagen)
- common: the latency dumped by “ceph osd perf” is not real ([issue#20749](#), [pr#16512](#), Pan Liu)
- common, tools: osdmaptool: show all the pg map to osds info ([pr#9419](#), song baisen)
- common: tracing: Fix handle leak in TracepointProvider ([pr#12652](#), Brad Hubbard)
- common: tracing: fix segv ([issue#18576](#), [pr#14304](#), Anjaneya Chagam)
- common: Update the error string when res\_nsearch() or res\_search() fails ([pr#15878](#), huanwen ren)
- common: use ref to avoid unnecessary memory copy ([issue#19107](#), [pr#13689](#), liuchang0812)
- common: use std::move() for better performance ([pr#16620](#), Xinying Song)
- common: xio: migrate atomic\_t to std::atomic<> ([pr#15230](#), Jesse Williamson)
- compressor: conditionalize on HAVE\_LZ4 ([pr#17174](#), Kefu Chai)
- compressor: fix Mutex::Locker used is not correct ([pr#13935](#), hechuang)
- compressor: zlib: fix plugin for non-Intel arches ([pr#14947](#), Dan Mick)
- core: auth: ‘ceph auth import -i’ overwrites caps, if caps are not specified ([issue#18932](#), [pr#13468](#), Vikhyat Umrao)

- core: auth: Remove unused function in AuthSessionHandler ([pr#16666](#), Luo Kexue)
- core: ceph: allow '--' with -i and -o for stdin/stdout ([pr#16359](#), Sage Weil)
- core: ceph-create-keys: Add connection timeouts ([pr#11995](#), Owen Synge)
- core: ceph-dencoder: Silence coverity CID 1412579 ([pr#15744](#), Brad Hubbard)
- core: ceph-detect-init: Add docker detection ([pr#13218](#), Guillaume Abrioux)
- core: ceph-disk: Adding retry loop in get\_partition\_dev() ([pr#14275](#), Erwan Velu)
- core: ceph-disk/ceph\_disk/main.py: fix calling of the bsdrc init scripts ([pr#14476](#), Willem Jan Withagen)
- core: ceph-disk/ceph\_disk/main.py: Replace ST\_ISBLK() test by is\_diskdevice() ([pr#15587](#), Willem Jan Withagen)
- core: ceph-disk: ceph-disk on FreeBSD should not use mpath-code ([pr#14837](#), Willem Jan Withagen)
- core: ceph-disk: dmcrypt activate must use the same cluster as prepare ([issue#17821](#), [pr#13573](#), Loic Dachary)
- core: ceph-disk: dmrypt cluster must default to ceph ([issue#20893](#), [pr#16776](#), Loic Dachary)
- core: ceph-disk: do not setup\_statedir on trigger ([issue#19941](#), [pr#15410](#), Loic Dachary)
- core: ceph-disk: enable directory backed OSD at boot time ([issue#19628](#), [pr#14546](#), Loic Dachary)
- core: ceph-disk: Fix getting wrong group name when -setgroup in bluestore ([issue#18955](#), [pr#13457](#), craigchi)
- core: ceph-disk: FreeBSD changes to get it working and passing tests ([pr#12086](#), Willem Jan Withagen)
- core: ceph-disk: implement prepare -no-locking ([pr#14728](#), Dan van der Ster, Loic Dachary)
- core: ceph\_disk/main.py: Allow FreeBSD zap a OSD disk ([pr#15642](#), Willem Jan Withagen)
- core: ceph-disk,osd: add support for crush device classes ([issue#19513](#), [pr#14436](#), Loic Dachary)
- core: ceph-disk: Populate mount options when running "list" ([issue#17331](#), [pr#14293](#), Brad Hubbard)
- core: ceph-disk: Reporting /sys directory in get\_partition\_dev() ([pr#14080](#), Erwan

Velu)

- core: ceph-disk: Revert “Revert “change get\_dmcrypt\_key test to support different cluster name”” ([pr#13600](#), Loic Dachary)
- core: ceph-disk: separate ceph-osd -check-needs-\* logs ([issue#19888](#), [pr#15016](#), Loic Dachary)
- core: ceph-disk: set the default systemd unit timeout to 3h ([issue#20229](#), [pr#15585](#), Loic Dachary)
- core: ceph-disk: support osd new ([pr#15432](#), Loic Dachary, Sage Weil)
- core: ceph-disk: Write 10M to all partitions before zapping ([issue#18962](#), [pr#13766](#), Wido den Hollander)
- core: ceph: do not throw TypeError on connection failure ([pr#13268](#), Kefu Chai)
- core: ceph.in: Fix couple of minor issues on the messages ([pr#12797](#), Dave Chen)
- core: ceph-objectstore-tool: do not populate snapmapper with missing clones ([issue#19943](#), [pr#15787](#), Sage Weil)
- core: ceph-osd: fix auto detect which objectstore is currently running ([issue#20865](#), [pr#16717](#), Yanhu Cao)
- core: ceph-osd: -flush-journal: sporadic segfaults on exit ([issue#18820](#), [pr#13311](#), Alexey Sheplyakov)
- core: client/SyntheticClient.cc: Fix warning in random\_walk ([issue#19445](#), [pr#14308](#), Brad Hubbard)
- core: cls/timeindex: clean up cls\_timeindex\_client.h|cc ([pr#13987](#), Shinobu Kinjo)
- core: common/options: remove mon\_warn\_osd\_usage\_min\_max\_delta from options.cc too ([pr#16488](#), Sage Weil)
- core: common/TrackedOp: allow dumping historic ops sorted by duration ([pr#14050](#), Piotr Dałek)
- core: compressor: add LZ4 support ([pr#15434](#), Haomai Wang)
- core: compressor: optimize header file dependency ([pr#15187](#), Brad Hubbard, Xiaowei Chen)
- core: Context: C\_ContextsBase: delete enclosed contexts in dtor ([issue#20432](#), [pr#16159](#), Kefu Chai)
- core: crush/CrushWrapper: chooseargs encoding fix ([pr#15984](#), Ilya Dryomov)
- core: crush/CrushWrapper: make get\_immediate\_parent[\_id] ignore per-class shadow hierarchy ([issue#20546](#), [pr#16221](#), Sage Weil)

- core: crush, mon: make jewel the lower bound for client/crush compat for new clusters ([pr#15370](#), Sage Weil)
- core: erasure-code: optimize header file dependency ([pr#15172](#), Brad Hubbard, Xiaowei Chen)
- core: erasure-code: Remove duplicate of isa-l files ([pr#15372](#), Ganesh Mahalingam)
- core: erasure-code: sync jerasure/gf-complete submodules ([pr#14424](#), Loic Dachary)
- core: filestore: migrate atomic\_t to std::atomic<> ([pr#15228](#), Jesse Williamson)
- core: Give requested scrub work a higher priority ([issue#15789](#), [pr#14488](#), David Zafman)
- core: global: start removing g\_ceph\_context ([pr#12149](#), Adam C. Emerson)
- core: HashIndex.cc: add compat.h for ENODATA ([pr#16697](#), Willem Jan Withagen)
- core: include/denc: add {encode,decode}\_nohead for denc\_traits<basic\_string> ([issue#18938](#), [pr#14099](#), Kefu Chai)
- core: include/mempool.h: fix Clangs complaint about types ([pr#13523](#), Willem Jan Withagen)
- core: include/types.h, introduce host\_to\_ceph\_errno ([pr#15496](#), Willem Jan Withagen)
- core: Install Pecan for FreeBSD ([pr#15610](#), Willem Jan Withagen)
- core: introduce (and fix) code to pass errno to other OSes ([pr#15495](#), Willem Jan Withagen)
- core: introduce DirectMessenger ([pr#14755](#), Casey Bodley, Matt Benjamin)
- core: kv/RocksDBStore: abort if rocksdb EIO, don't return incorrect result ([pr#15862](#), Haomai Wang)
- core: kv/RocksDBStore: use vector instead of VLA for holding slices ([pr#16615](#), Kefu Chai)
- core: libradosstriper: Initialize member variable m\_writeRc in WriteCompletionData ([pr#16780](#), amitkuma)
- core: luminous: Improve size scrub error handling and ignore system attrs in xattr checking ([issue#21051](#), [issue#18836](#), [issue#20243](#), [pr#17196](#), David Zafman)
- core: luminous: Include front/back interface names in OSD metadata ([issue#21048](#), [issue#20956](#), [pr#17193](#), John Spray)
- core: luminous: mon: bug in functon reweight\_by\_utilization ([issue#21079](#), [issue#20970](#), [pr#17198](#), xie xingguo)

- core: luminous: mon: “ceph osd crush rule rename” support ([pr#17260](#), xie xingguo)
- core: luminous: multisite: FAILED assert(prev\_iter != pos\_to\_prev.end()) in RGWMetaSyncShardCR::collect\_children() ([issue#21097](#), [issue#20906](#), [pr#17234](#), Casey Bodley)
- core: luminous: osd: osd\_scrub\_during\_recovery only considers primary, not replicas ([issue#18206](#), [issue#21077](#), [pr#17195](#), David Zafman)
- core: luminous: src/common/LogClient.cc: 310: FAILED assert(num\_unsent <= log\_queue.size()) ([issue#20965](#), [issue#18209](#), [pr#17197](#), Sage Weil)
- core: make the conversion from wire error to host OS work ([pr#15780](#), Willem Jan Withagen)
- core: Merge pull request #16755 from ivancich/wip-pull-new-dmclock ([pr#16922](#), Gregory Farnum)
- core: messages: default-initialize MOSDPGRecoveryDelete[Reply] members ([pr#16584](#), Greg Farnum)
- core: messages: Initialize members in MMDSTableRequest ([pr#16810](#), amitkuma)
- core: messages: Initialize member variables ([pr#16819](#), amitkuma)
- core: messages: Initialize member variables ([pr#16839](#), amitkuma)
- core: messages: Initializing member variable in MMDSCacheRejoin ([pr#16791](#), amitkuma)
- core: messages/MOSDOp: fix pg\_t decoding for version <7 decoding ([issue#19005](#), [pr#13537](#), Sage Weil)
- core: messages/MOSDPGTrim: add the missed HEAD\_VERSION AND COMPAT\_VERSION ([issue#18266](#), [pr#12517](#), huangjun)
- core: messages/MOSDPing.h: drop unused fields ([pr#15843](#), Piotr Dałek)
- core: messages/MOSDPing: initialize MOSDPing padding ([issue#20323](#), [pr#15714](#), Sage Weil)
- core: messages/MOSDSubOp: Make encode\_payload can be reentrant ([pr#12654](#), Haomai Wang)
- core: messages: remove compat cruft ([pr#14475](#), Sage Weil)
- core: mgr/MgrClient: do not attempt to access a global variable for config ([pr#16544](#), Jason Dillaman)
- core: mgr/MgrClient: use unique\_ptr for MgrClient::session ([issue#19097](#), [pr#13685](#), Kefu Chai)

- core,mgr: mgr/DaemonServer: stop spamming log with pg stats ([pr#15487](#), Sage Weil)
- core,mgr: mgr,librados: service map ([pr#15858](#), Yehuda Sadeh, John Spray, Sage Weil)
- core,mgr,mon: mgr,mon: enable/disable mgr modules via 'ceph mgr module ...' commands ([pr#15958](#), Sage Weil)
- core,mgr,mon: mon,mgr: tag some commands for ceph-mgr ([pr#13617](#), Sage Weil)
- core,mgr,mon: mon/PGMap: fix osd\_epoch update when removing osd\_stat ([issue#20208](#), [pr#15573](#), Sage Weil)
- core,mgr: mon/PGMap: slightly better debugging around pgmap updates ([pr#15820](#), Sage Weil)
- core,mgr,tests: qa: flush out monc's dropped msgs on msgr failure injection ([issue#20371](#), [pr#16484](#), Joao Eduardo Luis)
- core,mgr,tests: qa/suites/rados/rest: test restful mgr module ([pr#15604](#), Sage Weil)
- core: misc: SCA fixes ([pr#14426](#), Danny Al-Gaaf)
- core,mon: crush, mon: simplify device class manipulation commands ([pr#16388](#), xie xingguo)
- core: mon,mgr: fix "ceph osd df", add some tools to find untested commands ([issue#20256](#), [pr#15675](#), Greg Farnum)
- core: mon/MonClient: discard stray messages from non-acitve conns ([issue#19015](#), [pr#13656](#), Kefu Chai)
- core: mon/MonClient: don't return zero global\_id ([issue#19134](#), [pr#13853](#), "Yan, Zheng", Kefu Chai)
- core: mon/MonClient: hunt monitors in parallel ([issue#16091](#), [pr#11128](#), Steven Dieffenbach, Kefu Chai)
- core: mon/MonClient: persist global\_id across re-connecting ([issue#18968](#), [pr#13550](#), Kefu Chai)
- core: mon/MonClient: respect the priority in SRV RR ([issue#5249](#), [pr#15964](#), Kefu Chai)
- core,mon: mon/LogMonitor: 'log last' command ([pr#15497](#), Sage Weil)
- core: mon/MonmapMonitor: use \_\_func\_\_ instead of hard code function name ([pr#16037](#), Yanhu Cao)
- core,mon: mon/MgrStatMonitor: avoid dup health warnings during luminous upgrade ([issue#20435](#), [pr#15986](#), Sage Weil)

- core,mon: mon/MgrStatMonitor: keep mgrstat version ahead of pgmon ([issue#20219](#), [pr#15584](#), Sage Weil)
- core,mon: mon,osd: add crush\_version to OSDMap, and allow crush map updates to gate on crush\_version ([pr#15533](#), Sage Weil)
- core,mon: mon,osd: decouple creating pgs from pgmap ([pr#13999](#), Kefu Chai)
- core,mon: mon, osd: misc fixes ([pr#16078](#), xie xingguo)
- core,mon: mon/OSDMonitor: cancel mapping job from update\_from\_paxos ([issue#20067](#), [pr#15320](#), Sage Weil)
- core,mon: mon/OSDMonitor: make 'osd crush move ...' work on osds ([issue#18587](#), [pr#12981](#), Sage Weil)
- core,mon: mon/OSDMonitor: make snaps on tier pool should not be allowed ([pr#9348](#), Mingxin Liu)
- core,mon: mon/OSDMonitor: use up set instead of acting set in reweight\_by\_utilization ([pr#13802](#), Mingxin Liu)
- core,mon: mon,osd: new mechanism for managing full and nearfull OSDs for luminous ([pr#13615](#), Sage Weil)
- core,mon: mon/PGMap: call blocked requests ERR not WARN ([pr#15501](#), Sage Weil)
- core: mon,osd: add require\_min\_compat\_client setting to enforce and clarify client compatibility ([pr#14959](#), Sage Weil)
- core: mon,osd: luminous feature bits, require flags, upgrade gates ([pr#13278](#), Sage Weil)
- core: mon, osd: misc fixes and cleanups ([pr#16160](#), xie xingguo)
- core: mon, osd: misc fixes ([pr#16283](#), xie xingguo)
- core: mon/OSDMonitor: \_apply\_remap -> \_apply\_upmap; less code redundancy ([pr#15846](#), xie xingguo)
- core: mon/OSDMonitor: batch noup/noin osds support ([pr#15725](#), xie xingguo)
- core: mon/OSDMonitor: batch OSDs nodown/noout support ([pr#15381](#), xie xingguo)
- core: mon/OSDMonitor: change info in 'osd failed' messages ([pr#15321](#), Sage Weil)
- core: mon/OSDMonitor: do not allow crush device classes until luminous ([pr#16188](#), Sage Weil)
- core: mon/OSDMonitor: fixup sortbitwise flag warning ([pr#12682](#), huanwen ren)
- core: mon/OSDMonitor: make mapping job behave if mon\_osd\_prime\_pg\_temp = false

([issue#19020](#), [pr#13574](#), Sage Weil)

- core: mon/OSDMonitor: osd crush set-device-class ([issue#19307](#), [pr#14039](#), Loic Dachary)
- core: mon/OSDMonitor: set last\_force\_op\_resend on overlay pool too ([issue#18366](#), [pr#12712](#), Sage Weil)
- core: mon/OSDMonitor: should propose osdmap update when cluster addr changed ([pr#11065](#), Mingxin Liu)
- core: mon/OSDMonitor: skip prime\_pg\_temp if mapping is prior to osdmap ([pr#14826](#), Kefu Chai)
- core: mon,osd/OSDMap: a couple pg-upmap fixes ([pr#15319](#), Sage Weil)
- core: mon/PGMap: factor mon\_osd\_full\_ratio into MAX AVAIL calc ([issue#18522](#), [pr#12923](#), Sage Weil)
- core: mon/PGMonitor: fix wrongly report "pg stuck in inactive" ([pr#14391](#), Mingxin Liu)
- core,mon,rbd: mon,osd: new rbd-based cephx cap profiles ([pr#15991](#), Jason Dillaman)
- core: msg/async/AsyncConnection: keepalive objecter ping connection to avoid timeout ([pr#14009](#), Haomai Wang)
- core: msg/async/AsyncConnection: socket's fd can be zero, avoid false assert ([pr#13080](#), Haomai Wang)
- core: msg/async: avoid requeue racing with handle\_write ([issue#20093](#), [pr#15324](#), Haomai Wang)
- core: msg/async/dpdk: fix compile errors ([pr#12698](#), Haomai Wang)
- core: msg/async: fix deleted\_conn is out of sync with conns ([issue#20230](#), [pr#15645](#), Haomai Wang)
- core: msg/async: fix the bug of inaccurate calculation of l\_msgr\_send\_bytes ([pr#16526](#), Jin Cai)
- core: msg/async/rdma: add log to show correct destruct queuepair ([pr#13412](#), Haomai Wang)
- core: msg/async/rdma: add perf counters to RDMA backend ([pr#13484](#), Haomai Wang)
- core: msg/async/rdma: destroy QueuePair if needed ([pr#13810](#), Haomai Wang)
- core: msg/async/rdma: don't need to delete event when tcp connection isn't ... ([pr#13528](#), Haomai Wang)

- core: msg/async/rdma: fix ceph\_clock\_now calls ([pr#12711](#), Haomai Wang)
- core: msg/async/rdma: fix potential racing connection usage ([pr#13738](#), Haomai Wang)
- core: msg/async/rdma: make Infiniband can be forkable ([pr#13525](#), Haomai Wang)
- core: msg/async/rdm: fix leak when existing failure in ip network ([pr#13435](#), Haomai Wang)
- core: msg/async: set thread name for msgr worker ([pr#13699](#), Haomai Wang)
- core: msg/async/Stack.cc: use of pthread\_setname\_np() needs compat.h ([pr#13825](#), Willem Jan Withagen)
- core: msg/async: support IPv6 QoS ([issue#18887](#), [issue#18928](#), [pr#13418](#), Robin H. Johnson)
- core: msg/simple: fix missing unlock when already bind ([pr#13267](#), Haomai Wang)
- core: msg/simple/Pipe: the returned value for do\_recv unequal to zero ([pr#10272](#), zhang.zezhu)
- core: objclass: modify omap\_get\_{keys,vals} api ([pr#16667](#), Yehuda Sadeh, Casey Bodley)
- core: objclass-sdk: use namespace ceph for bufferlist ([pr#15581](#), Neha Ojha)
- core: os/bluestore: do not use nullptr to calc the size of bluestore\_pextent\_t ([pr#14030](#), Kefu Chai)
- core: os/bluestore rm unused variable in aio\_read() ([pr#13530](#), tangwenjun)
- core: os/bluestore: silence gcc warning ([pr#14028](#), Kefu Chai)
- core: osdc: clean up osd\_command/start\_mon\_command interfaces ([pr#13727](#), John Spray)
- core: osdc/Objecter: fix possible OSDSession leak on wrong connection ([pr#13365](#), xie xingguo)
- core: osdc/Objecter: resend pg commands on interval change ([issue#18358](#), [pr#12869](#), Samuel Just)
- core: osdc/Objecter: respect epoch barrier in \_op\_submit() ([issue#19396](#), [pr#14190](#), Ilya Dryomov)
- core: osd/: don't leak context for Blessed\*Context or RecoveryQueueAsync ([issue#18809](#), [pr#13342](#), Samuel Just)
- core: OSD: drop parameter t from merge\_log() ([pr#13923](#), xie xingguo)

- core: osd/ECBackend: cleanup for unnecessary copy with pg\_stat\_t ([pr#12564](#), Yunchuan Wen)
- core: osd/ECBackend: drop duplicated pending\_commit field from << operator ([pr#13665](#), xie xingguo)
- core: osd/ECBackend: only need check missing\_loc when doing recovery ([pr#12526](#), huangjun)
- core: osd/ECBackend: remove unused variable "ReadCB" ([pr#12543](#), huangjun)
- core: osd/ECTransaction: cleanup the redundant check which works in overwrite IO context ([pr#15765](#), tang.jin)
- core: osd/ECTransaction: only read partial stripes when below \*original\* object size ([issue#19882](#), [pr#15712](#), Sage Weil)
- core: osd/filestore: Revert "os/filestore: move ondisk in front ([issue#20524](#), [pr#16156](#), Kefu Chai)
- core: osd,librados: add manifest, redirect ([pr#15325](#), Sage Weil)
- core: osd,librados: cmpext support ([pr#14715](#), Zhengyong Wang, David Disseldorp, Mike Christie)
- core: osd,librados: remove clone\_range and associated multi-object cruft ([pr#13008](#), Samuel Just)
- core: osd, messages/MOSDPing: bunch of fixes related to ping inflation ([pr#15727](#), Piotr Dałek)
- core: osd/mon/mds: fix config set tell command ([issue#20803](#), [pr#16700](#), John Spray)
- core: osd,mon: misc full fixes and cleanups ([pr#13968](#), David Zafman)
- core: osd/OpRequest: dump both name and addr for the client op ([pr#12691](#), runsisi)
- core: osd/OSD: bump up current version; conditionally encoding manifest into oi ([pr#15687](#), xie xingguo)
- core: osd/osd\_internal\_types: wake snaptrimmer on put\_read lock, too ([issue#19131](#), [pr#13755](#), Sage Weil)
- core: osd/OSDMap: bump encoding version for require\_min\_compat\_client ([pr#15046](#), "Yan, Zheng")
- core: osd/OSDMap: Change \*pg\_to\_\* to return void ([pr#15684](#), Brad Hubbard)
- core: osd/OSDMap: don't set weight to IN when OSD is destroyed ([issue#19119](#), [pr#13730](#), Ilya Dryomov)

- core: osd/OSDMap: hide require\_osd and sortbitwise flags ([pr#14440](#), Sage Weil)
- core: osd/OSDMap: improve upmap calculation ([issue#19818](#), [pr#14902](#), Sage Weil)
- core: osd/OSDMap: Uncomment code to enable private default constructors ([pr#12597](#), Brad Hubbard)
- core: osd/OSD: tolerate any 'set-device-class' error on OSD startup ([pr#16812](#), xie xingguo)
- core: osd/osd\_type: Fix logging output ([pr#12778](#), Brad Hubbard)
- core: osd/osd\_types: Move comment to more relevant position ([pr#12779](#), Brad Hubbard)
- core: osd/osd\_types: print notify-ack op properly ([pr#12585](#), Sage Weil)
- core: osd/PG: add new have\_unfound() function in MissingLoc ([pr#12668](#), huangjun)
- core: osd/PG: Add two new mClock implementations of the PG sharded operator queue ([pr#14997](#), J. Eric Ivancich)
- core: osd/PG.cc: Optimistic estimation on PG.last\_active ([pr#14799](#), Xiaoxi Chen)
- core: osd/PG.cc: unify the call of checking whether lock is held ([pr#15013](#), Jin Cai)
- core: osd/PG: check the connection first in fulfill\_log ([pr#12579](#), huangjun)
- core: osd/PG: conditionally retry on receiving pg-notify when Primary is Incomplete ([pr#13942](#), xie xingguo)
- core: osd/PG: drop pre-firefly compat\_mode for choose\_acting ([pr#15057](#), Sage Weil)
- core: osd/PG: fix lost unfound + delete when there are no missing objects ([issue#20904](#), [pr#16809](#), Josh Durgin)
- core: osd/PG: fix possible overflow on unfound objects ([pr#12669](#), huangjun)
- core: osd/PG: fix warning so we discard\_event() on a no-op state change ([pr#16655](#), Sage Weil)
- core: osd/PG: ignore CancelRecovery in NotRecovering ([issue#20804](#), [pr#16638](#), Sage Weil)
- core: osd/PGLog: avoid infinite loop if missing version is corrupted ([pr#16798](#), Josh Durgin)
- core: osd/PGLog: fix inaccurate missing assert ([issue#20753](#), [pr#16539](#), Josh Durgin)

- core: osd/PGLog: fix index for parent and child log on split ([issue#18975](#), [pr#13493](#), Sage Weil)
- core: osd/pglog: remove loop through empty collection ([pr#15121](#), J. Eric Ivancich)
- core: osd/PGLog: skip ERROR entires in \_merge\_object\_divergent\_entries ([issue#20843](#), [pr#16675](#), Jeegn Chen)
- core: osd/PG: make non-empty PastIntervals non-fatal ([issue#20167](#), [pr#15639](#), Sage Weil)
- core: osd/PG: only correct filestore collection bits on load ([issue#19541](#), [pr#14397](#), Sage Weil)
- core: osd/PG: publish PG stats when backfill-related states change ([issue#18369](#), [pr#12727](#), Sage Weil)
- core: osd/PG: reset the missing set when restarting backfill ([issue#19191](#), [pr#14053](#), Josh Durgin)
- core: osd/PG: restrict want\_acting to up+acting on recovery completion ([issue#18929](#), [pr#13420](#), Sage Weil)
- core: osd/PG: set clean when last\_epoch\_clean is updated ([issue#19023](#), [pr#15555](#), Samuel Just)
- core: osd/PG: simplify the logic of backfill\_targets checking ([pr#12519](#), huangjun)
- core: osd/PG: some minor cleanups ([pr#14133](#), runsisi)
- core: osd/PrimaryLogPG: clear oi from trim\_object() ([issue#19947](#), [pr#15519](#), Sage Weil)
- core: osd/PrimaryLogPG: do not call on\_shutdown() if (pg.deleting) ([issue#19902](#), [pr#15040](#), Kefu Chai)
- core: osd/PrimaryLogPG: do not expect FULL\_TRY ops to get resent ([issue#19430](#), [pr#14255](#), Sage Weil)
- core: osd/PrimaryLogPG::failed\_push: update missing as well ([issue#18165](#), [pr#12888](#), Samuel Just)
- core: osd/PrimaryLogPG: fix oi reset during trim\_object ([issue#19947](#), [pr#15696](#), Sage Weil)
- core: osd/PrimaryLogPG: fix recovering hang when have unfound objects ([pr#16558](#), huangjun)
- core: osd/PrimaryLogPG: optimal pick\_newest\_available ([pr#12695](#), huangjun)

- core: osd/PrimaryLogPG: record prior\_version for DELETE events ([issue#20274](#), [pr#15649](#), Sage Weil)
- core: osd/PrimaryLogPG: remove duplicated code ([pr#13894](#), Jianpeng Ma)
- core: osd/PrimaryLogPG: set return value if sparse read failed ([pr#14093](#), huangjun)
- core: osd/PrimaryLogPG: skip deleted missing objects in pg[n]ls ([issue#20739](#), [pr#16490](#), Josh Durgin)
- core: osd/PrimaryLogPG solve cache tier osd high memory consumption ([issue#20464](#), [pr#16011](#), Peng Xie)
- core: osd/PrimaryLogPG::try\_lock\_for\_read: give up if missing ([issue#18583](#), [pr#13087](#), Samuel Just)
- core: osd/PrimaryLogPG: unify the access to primary pg ([pr#12527](#), huangjun)
- core: osd/PrimayLogPG: update modified range to include the whole object size for write\_full op ([pr#15021](#), runsisi)
- core: osd/ReplicatedBackend: clear pull source once we are done with it ([issue#19076](#), [pr#13879](#), Samuel Just)
- core: osd/ReplicatedBackend: remove MOSDSubOp cruft from repop\_applied ([pr#14358](#), Jianpeng Ma)
- core: osd/ReplicatedBackend: reset thread heartbeat after every omap entry ... ([issue#20375](#), [pr#15823](#), Josh Durgin)
- core: osd/ReplicatedBackend: take read locks for clone sources during recovery ([issue#17831](#), [pr#12844](#), Samuel Just)
- core: os/filestore: call committed\_thru when no journal entries are replayed ([pr#15781](#), Kuan-Kai Chiu)
- core: os/filestore: debug which omap keys are set ([issue#19067](#), [pr#13671](#), Sage Weil)
- core: os/filestore: do not free event if not added ([pr#16235](#), Kefu Chai)
- core: os/filestore/FileJournal: bufferlist rebuild ([pr#13980](#), Jianpeng Ma)
- core: os/filestore/FileJournal: FileJournal::open() close journal file before return error ([issue#20504](#), [pr#16120](#), Yang Honggang)
- core: os/filestore/FileStore.cc: remove a redundant judgement when get max latency ([pr#15961](#), Jianpeng Ma)
- core: os/filestore/FileStore.cc: remove unneeded loop ([pr#12177](#), Li Wang)

- core: os/filestore: fix clang static check warn “use-after-free” ([pr#12581](#), liuchang0812)
- core: os/filestore: fix infinit loops in fiemap() ([pr#14367](#), Ning Yao)
- core: os/filestore: handle error returned from write\_fd() ([pr#10146](#), yonghengdexin735)
- core: os/filestore/HashIndex: be loud about splits ([issue#18235](#), [pr#12421](#), Dan van der Ster)
- core: os/filestore/JournalingObjectStore cleanup ([pr#12528](#), Li Wang)
- core: os/filestore: require experimental flag for btrfs ([pr#16086](#), Sage Weil)
- core: os/filestore: version will be uninitialized variable if store\_version doesn't exist ([pr#12582](#), liuchang0812)
- core: os/fs/FS.cc: remove the redundant code ([pr#14362](#), Jianpeng Ma)
- core: os/FuseStore: include <functional> header in src/os/FuseStore.h for gcc 7.x ([pr#13454](#), Jos Collin)
- core, performance: common/config\_opts: improve rdma buffer size to 128k ([pr#13510](#), Haomai Wang)
- core, performance: common/TrackedOp: various cleanups and optimizations ([pr#12537](#), Sage Weil)
- core, performance: kv/RocksDBStore: Table options for indexing and filtering ([pr#16450](#), Mark Nelson)
- core, performance: mon,osd: explicitly remap some pgs ([pr#13984](#), Sage Weil)
- core, performance: msg/async: avoid lossy connection sending ack message ([pr#13700](#), Haomai Wang)
- core, performance: msg/async/rdma: cleanup ([pr#13509](#), Haomai Wang)
- core, performance: msg/async/rdma: refactor tx handle flow to get rid of locks ([pr#13680](#), Haomai Wang)
- core, performance: msg/async: reduce write\_lock contention ([pr#15092](#), Haomai Wang)
- core, performance: osd/ECBackend: Send write message to peers first, then do local write ([pr#12522](#), huangjun)
- core, performance: osd/OSD.h: requeue the scrub job with higher priority to shorten the blocking time of related requests ([pr#15552](#), Jin Cai)
- core, performance: osd, os: reduce fiemap burden ([pr#14640](#), Piotr Dałek)

- core, performance: osd/pg: bound the portion of the log we request in GetLog::GetLog() ([pr#12233](#), Jie Wang)
- core, performance: osd/PG: make prioritized recovery possible ([pr#13723](#), Piotr Dałek)
- core, performance: os/filestore: avoid unnecessary copy in filestore::\_do\_transaction ([pr#12578](#), Yunchuan Wen)
- core, performance: os/filestore/HashIndex: randomize split threshold by a configurable amount ([issue#15835](#), [pr#15689](#), Josh Durgin)
- core, performance: os/filestore: queue ondisk completion before apply work ([pr#13918](#), Pan Liu)
- core, performance: os/filestore: use new sleep strategy when io\_submit gets EAGAIN ([pr#14860](#), Pan Liu)
- core, performance: os/kstore: Added rocksdb bloom filter settings ([pr#13053](#), Ted-Chang)
- core, performance: src/OSD: add more useful perf counters for performance tuning ([pr#15915](#), Pan Liu)
- core: PGLog: store extra duplicate ops beyond the normal log entries ([pr#16172](#), Josh Durgin, J. Eric Ivancich)
- core: Prefix /proc/ with FreeBSD emulation ([pr#14290](#), Willem Jan Withagen)
- core: PrimaryLogPG: don't update digests for objects with mismatched names ([issue#18409](#), [pr#12788](#), Samuel Just)
- core: print more information when run ceph-osd cmd with 'check options' ([pr#16678](#), mychoxin)
- core: qa: do not restrict valgrind runs to centos ([issue#18126](#), [pr#15389](#), Greg Farnum)
- core, rbd: mon, osd: do not create rbd pool by default ([pr#15894](#), Greg Farnum, Sage Weil, David Zafman)
- core: ReplicatedBackend: don't queue Context outside of ObjectStore with obc ([issue#18927](#), [pr#13569](#), Samuel Just)
- core: Revert "PrimaryLogPG::failed\_push: update missing as well" ([issue#18624](#), [pr#13090](#), David Zafman)
- core, rgw: misc: SCA and Coverity Fixes ([pr#13208](#), Danny Al-Gaaf)
- core, rgw: qa: Removed all 'default\_idle\_timeout' due to chnage in rgw task ([pr#15420](#), Yuri Weinstein)

- core,rgw,tests: qa/rgw\_snaps: move default\_idle\_timeout config under the client ([issue#20128](#), [pr#15400](#), Yehuda Sadeh)
- core,rgw,tests: qa/suits/rados/basic/tasks/rgw\_snaps: wait for pools to be created ([pr#16509](#), Sage Weil)
- core: rocksdb: sync with upstream ([issue#18464](#), [pr#13306](#), Kefu Chai)
- core: src/ceph.in: Use env(CEPH\_DEV) to suppress noise from ceph ([pr#14746](#), Willem Jan Withagen)
- core: src/vstart.sh: kill dead upmap option ([pr#15848](#), xie xingguo)
- core:" Stringify needs access to << before reference" src/include/stringify.h ([pr#16334](#), Willem Jan Withagen)
- core: test, osd: fix some coverity issues ([pr#13293](#), liuchang0812)
- core: test/pybind/test\_rados.py: tolerate TimedOut in test\_ping\_monitor ([issue#18529](#), [pr#12934](#), Samuel Just)
- core,tests: ceph-disk: sensible default for block.db ([pr#15576](#), Loic Dachary)
- core,tests: ceph-disk/tests: Certain partition types do not work on FreeBSD ([pr#13560](#), Willem Jan Withagen)
- core,tests: ceph-disk/tests/test\_main.py: FreeBSD does not do multipath ([pr#13847](#), Willem Jan Withagen)
- core,tests: ceph\_test\_librados\_api\_misc: fix stupid LibRadosMiscConnectFailure.ConnectFailure test ([issue#15368](#), [pr#14261](#), Sage Weil)
- core,tests: ceph\_test\_rados\_api\_misc: avoid livelock from PoolCreationRace ([pr#13565](#), Sage Weil)
- core,tests: ceph\_test\_rados\_api\_misc: Fix trivial memory leak ([pr#12680](#), Brad Hubbard)
- core,tests: ceph\_test\_rados\_api: wait for snap trim on ENOENT during cleanup ([issue#19948](#), [pr#15638](#), Sage Weil)
- core,tests: ceph\_test\_rados\_api\_watch\_notify: flush after unwatch ([issue#20105](#), [pr#16402](#), Sage Weil)
- core,tests: ceph\_test\_rados\_api\_watch\_notify: make LibRadosWatchNotify.Watch3Timeout tolerate thrashing ([issue#19433](#), [pr#14899](#), Sage Weil)
- core,tests: ceph\_test\_rados: max\_stride\_size must be more than min\_stride\_size ([issue#20775](#), [pr#16590](#), Lianne Wang)
- core,tests: c\_write\_operations.cc: Fix trivial memory leak ([pr#12663](#), Brad

Hubbard)

- core,tests: do all valgrind runs on centos ([issue#20360](#), [issue#18126](#), [pr#16046](#), Sage Weil)
- core,tests: os: allow 'osd objectstore = random' to pick either filestore or bluestore ([pr#13754](#), Sage Weil)
- core,tests: qa: avoid map-gap tests for k=2 m=1 ([issue#20844](#), [pr#16789](#), Sage Weil)
- core,tests: qa: move ceph-helpers-based make check tests to qa/standalone; run via teuthology ([pr#16513](#), Sage Weil)
- core,tests: qa/objectstore/filestore-btrfs: test btrfs on trusty only ([issue#20169](#), [pr#15814](#), Sage Weil)
- core,tests: qa/objectstore: test bluestore with aggressive compression ([pr#14623](#), Sage Weil)
- core,tests: qa/rados/upgrade/jewel-x-singleton: run luminous.yaml at the end ([pr#13378](#), Sage Weil)
- core,tests: qa: stop testing btrfs ([issue#20169](#), [pr#16044](#), Sage Weil)
- core,tests: qa/suites/powercycle/osd/tasks/radosbench: consume less space ([issue#20302](#), [pr#15821](#), Sage Weil)
- core,tests: qa/suites/rados: at-end: ignore PG\_{AVAILABILITY,DEGRADED} ([issue#20693](#), [pr#16575](#), Sage Weil)
- core,tests: qa/suites/rados/\*/at-end: wait for healthy before scrubbing ([pr#15245](#), Sage Weil)
- core,tests: qa/suites/rados/basic: set low omap limit for rgw workload ([pr#13071](#), Sage Weil)
- core,tests: qa/suites/rados/basic/tasks/rados\_python: POOL\_APP\_NOT\_ENABLED ([pr#16827](#), Sage Weil)
- core,tests: qa/suites/rados/mgr/tasks/failover: whitelist ([pr#16795](#), Sage Weil)
- core,tests: qa/suites/rados/singleton/all/reg11184: whitelist health warnings ([pr#16306](#), Sage Weil)
- core,tests: qa/suites/rados/singleton-nomsg/health-warnings: behave on ext4 ([issue#20043](#), [pr#15207](#), Sage Weil)
- core,tests: qa/suites/rados: temporarily remove scrub\_test from basic/ until post-luminous ([issue#19935](#), [pr#15202](#), Sage Weil)
- core,tests: qa/suites/rados/thrash/workload/\*: enable rados.py cache tiering ops

([issue#11793](#), [pr#16244](#), Sage Weil)

- core,tests: qa/suites/upgrade/kraken-x: enable experimental for bluestore ([pr#15359](#), Sage Weil)
- core,tests: qa/tasks/ceph: enable rbd on rbd pool ([pr#16794](#), Sage Weil)
- core,tests: qa/tasks/ceph\_manager: get osds all in after thrashing ([pr#15784](#), Sage Weil)
- core,tests: qa/tasks/ceph\_manager: wait for osd to start after objectstore-tool sequence ([issue#20705](#), [pr#16454](#), Sage Weil)
- core,tests: qa/tasks/ceph\_manager: wait longer for pg stats to flush ([pr#16322](#), Sage Weil)
- core,tests: qa/tasks/ceph: osd\_scrub\_pgs: reissue scrub requests in loop ([issue#20326](#), [pr#15747](#), Sage Weil)
- core,tests: qa/tasks/ceph.py: no osd id to 'osd create' command ([issue#20548](#), [pr#16233](#), Sage Weil)
- core,tests: qa/tasks/ceph.py: tolerate active+clean+something ([pr#15717](#), Sage Weil)
- core,tests: qa/tasks/ceph: simplify ceph deployment slightly ([pr#15853](#), Sage Weil)
- core,tests: qa/tasks/ceph: wait for mgr to activate and pg stats to flush in health() ([issue#20744](#), [pr#16514](#), Sage Weil)
- core,tests: qa/tasks/dump\_stuck: fix dump\_stuck test bug ([pr#16559](#), huangjun)
- core,tests: qa/tasks/dump\_stuck: fix for active+clean+remapped ([issue#20431](#), [pr#15955](#), Sage Weil)
- core,tests: qa/tasks/radosbench: longer timeout ([pr#16213](#), Sage Weil)
- core,tests: qa/workunits/cephtool/test.sh: add sudo for daemon compact ([pr#16500](#), Sage Weil)
- core,tests: qa/workunits/cephtool/test.sh: fix osd full health detail grep ([issue#20187](#), [pr#15494](#), Sage Weil)
- core,tests: qa/workunits/rados/test\_health\_warning: misc fixes ([issue#19990](#), [pr#15201](#), Sage Weil)
- core,tests: qa/workunits/rest: use unique pool names for cephfs test ([pr#13188](#), Sage Weil)
- core,tests: Revert "qa: do not restrict valgrind runs to centos" ([issue#20360](#), [pr#15791](#), Sage Weil)

- core,tests: test: add separate ceph-helpers-based smoke test ([pr#16572](#), Sage Weil)
- core,tests: test/librados/cmd.cc: Fix trivial memory leaks ([pr#12671](#), Brad Hubbard)
- core,tests: test/librados/c\_read\_operations.cc: Fix trivial memory leak ([pr#12656](#), Brad Hubbard)
- core,tests: test/librados/c\_read\_operations.cc: Fix valgrind errors ([issue#18354](#), [pr#12657](#), Brad Hubbard)
- core,tests: test/librados: Silence Coverity memory leak warnings ([pr#12442](#), Brad Hubbard, Samuel Just)
- core,tests: test/librados/snapshots.cc: Fix memory leak ([pr#12690](#), Brad Hubbard)
- core,tests: test/librados/tier.cc: Fix valgrind errors ([issue#18360](#), [pr#12705](#), Brad Hubbard)
- core,tests: test/osd/TestRados.cc: run set-redirect test after finishing setup ([issue#20114](#), [pr#15385](#), Myoungwon Oh)
- core,tests: test\_rados\_watch\_notify: Fix trivial memory leaks ([pr#12713](#), Brad Hubbard)
- core,tests,tools: Fixes: <http://tracker.ceph.com/issues/18533> ([pr#13423](#), Samuel Just, David Zafman)
- core,tests: upgrade/jewel-x: a few fixes ([pr#16830](#), Sage Weil)
- core: throttle: Minimal destructor fix for Luminous ([pr#16661](#), Adam C. Emerson)
- core,tools: ceph: perfcounter priorities and daemonperf updates to use them ([pr#14793](#), Sage Weil, Dan Mick)
- core,tools: kv: move 'bluestore-kv' hackery out of KeyValueDB into ceph-kvstore-tool ([issue#19778](#), [pr#14895](#), Sage Weil)
- core,tools: osdmaptool: require -upmap-save before modifying input osdmap ([pr#15247](#), Sage Weil)
- core: vstart.sh: start mgr after mon, before osds ([pr#16613](#), Sage Weil)
- core: Wip 20985 divergent handling luminous ([issue#20985](#), [pr#17001](#), Greg Farnum)
- create the ceph-volume and ceph-volume-systemd man pages ([pr#17158](#), Alfredo Deza)
- crush: a couple of weight-set fixes ([pr#16623](#), xie xingguo)
- crush: add devices class that rules can use as a filter ([issue#18943](#), [pr#13444](#), Loic Dachary)

- crush: add -dump to crushtool ([pr#13726](#), Loic Dachary)
- crush: add missing tunable in tests ([pr#15412](#), Loic Dachary)
- crush: allow uniform buckets with no items ([pr#13521](#), Loic Dachary)
- crush: API documentation ([pr#13205](#), Loic Dachary)
- crush: bucket: crush\_add\_uniform\_bucket\_item should check for uniformity ([pr#14208](#), Sahid Orentino Ferdjaoui)
- crush: builder: clean the arguments of crush\_reweight\* methods ([pr#14110](#), Sahid Orentino Ferdjaoui)
- crush: builder: creating crush map with optimal configurations ([pr#14209](#), Sahid Orentino Ferdjaoui)
- crush: builder: legacy has chooseleaf\_stable = 0 ([pr#14695](#), Loic Dachary)
- crush: crush\_init\_workspace starts with struct crush\_work ([pr#14696](#), Loic Dachary)
- crush: detect and (usually) fix ruleset != rule id ([pr#13683](#), Sage Weil)
- crush: document tunables and rule step set ([pr#13722](#), Loic Dachary)
- crush: do is\_out test only if we do not collide ([pr#13326](#), xie xingguo)
- crush: encode can override weights with weight set ([issue#19836](#), [pr#15002](#), Loic Dachary)
- crush: enforce buckets-before-rules rule ([pr#16453](#), Sage Weil)
- crush: fix CrushCompiler won't compile maps with empty shadow tree ([pr#17228](#), xie xingguo)
- crush: fix dprintk compilation ([pr#13424](#), Loic Dachary)
- crush: force rebuilding shadow hierarchy after swapping buckets ([pr#17229](#), xie xingguo)
- crush: misc changes/fixes for device classes ([issue#20845](#), [pr#16805](#), Kefu Chai, xie xingguo, Sage Weil)
- crush: more class fixes ([pr#16837](#), xie xingguo)
- crush: only encode class info if SERVER\_LUMINOUS ([issue#19361](#), [pr#14131](#), Sage Weil)
- crush: optimize header file dependency ([pr#9307](#), Xiaowei Chen)
- crush: silence warning from -Woverflow ([pr#16329](#), Jos Collin)

- crush: s/ruleset/id/ in decompiled output; prevent compilation when ruleset != id ([pr#16400](#), Sage Weil)
- crush: update choose\_args when items are added/removed ([pr#15311](#), Loic Dachary)
- crush: update documentation for negative choose step ([pr#14970](#), Loic Dachary)
- crush: various weight-set fixes ([pr#17214](#), xie xingguo)
- crush: verify weights is influenced by the number of replicas ([issue#15653](#), [pr#13083](#), Adam C. Emerson, Loic Dachary)
- crush: weight\_set and id remapping ([issue#15653](#), [pr#14486](#), Loic Dachary)
- crush: when osd\_location\_hook does not exist, we should exit error ([pr#12961](#), song baisen)
- doc: 12.1.0/release notes 2 ([pr#15627](#), Abhishek Lekshmanan)
- doc: 12.1.1 & 12.1.2 release notes ([pr#16377](#), Abhishek Lekshmanan)
- doc: add 0.94.10 and hammer EOL to releases.rst ([pr#13069](#), Nathan Cutler)
- doc: add 12.0.1 release notes ([pr#14106](#), Abhishek Lekshmanan)
- doc: Add amitkumar50 affiliation to .organizationmap ([pr#16475](#), Amit Kumar)
- doc add ceph-volume and ceph-volume-systemd man pages to CMakeLists file ([pr#17170](#), Alfredo Deza)
- doc: add changelog for v0.94.10 ([pr#13572](#), Abhishek Lekshmanan)
- doc: add changelog for v10.2.6 Jewel release ([pr#13839](#), Abhishek Lekshmanan)
- doc: add changelog for v10.2.7 ([pr#14441](#), Abhishek Lekshmanan)
- doc: add descriptions for mon/mgr options ([pr#15032](#), Kefu Chai)
- doc: add doc requirements on PR submitters ([pr#16394](#), John Spray)
- doc: added mgr caps to manual deployment documentation ([pr#16660](#), Nick Erdmann)
- doc: add FreeBSD manual install ([pr#14941](#), Willem Jan Withagen)
- doc: add instructions for replacing an OSD ([pr#16314](#), Kefu Chai)
- doc: add new cn ceph mirror to doc and mirroring ([pr#15089](#), Shengjing Zhu)
- doc: add optional argument for build-doc ([pr#14058](#), Kefu Chai)
- doc: add rados xattr commands to manpage ([pr#15362](#), Andreas Gerstmayr)
- doc: add rbd new trash cli and cleanups in release-notes.rst ([issue#20702](#),

- pr#16498, songweibin)
- doc: add README to dmclock subdir to inform developers it's a git subtree (pr#15386, J. Eric Ivancich)
- doc: add RGW ldap auth documentation (pr#14339, Harald Klein)
- doc: add some undocumented options to rbd-nbd (pr#14134, wangzhengyong)
- doc: add verbiage to rbdmap manpage (issue#18262, pr#12509, Nathan Cutler)
- doc: Add Zabbix ceph-mgr plugin to PendingReleaseNotes (pr#16412, Wido den Hollander)
- doc: AUTHORS: update CephFS PTL (pr#16399, Patrick Donnelly)
- doc: AUTHORS: update tech leads (pr#14350, Patrick Donnelly)
- doc: AUTHORS: update with release manager, backport team (pr#15391, Sage Weil)
- doc: build/install-deps.sh: Add sphinx package for building docs on FreeBSD (pr#13223, Willem Jan Withagen)
- doc: ceph-disk: use '-' for feeding ceph cli with stdin (pr#16362, Kefu Chai)
- doc: change osd\_op\_thread\_timeout default value to 15 (pr#14199, Andreas Gerstmayr)
- doc: Change the default values of some OSD options (issue#20199, pr#15566, Bara Ancincova)
- doc: clarify "ceph quorum" syntax (issue#17802, pr#11787, Nathan Cutler)
- doc: clarify SubmittingPatches.rst (pr#12988, Nathan Cutler)
- doc: clarify that "ms bind ipv6" disables IPv4 (pr#13317, Ken Dreyer)
- doc: clarify the path restriction mds cap example (pr#12993, John Spray)
- doc: common/options.cc: document bluestore config options (pr#16489, Sage Weil)
- doc: correct and improve add user capability section (pr#14055, Chu, Hua-Rong)
- doc: correct arguments for ceph tell osd.N bench (pr#14462, Patrick Dinnen)
- doc: Correcting the remove bucket example and adding bucket link/unlink examples (pr#12460, Uday Mullangi)
- doc: correct S3 lifecycle support explain (issue#18459, pr#12827, liuchang0812)
- doc: correct the quota section (issue#19397, pr#14122, Chu, Hua-Rong)
- doc: crush: API documentation fixes (pr#13589, Loic Dachary)

- doc: crush typo in algorithm description ([pr#13661](#), Loic Dachary)
- doc: deletes duplicated word and clarifies an example ([pr#13746](#), Tahia Khan)
- doc: describe CephFS max\_file\_size ([pr#15287](#), Ken Dreyer)
- doc: describe mark\_events logging available via the OSD's OpTracker ([pr#15095](#), Greg Farnum)
- doc: Describe mClock's use within Ceph in great detail ([pr#16707](#), J. Eric Ivancich)
- doc: dev add a note about ccache ([pr#14478](#), Abhishek Lekshmanan)
- doc: dev: add notes on PR make check validation test ([pr#16079](#), Nathan Cutler)
- doc: dev guide: how to run s3-tests locally against vstart ([pr#14508](#), Nathan Cutler, Abhishek Lekshmanan)
- doc: dev improve the s3tests doc to reflect current scripts ([pr#15180](#), Abhishek Lekshmanan)
- doc: doc/cephfs: mention RADOS object size limit ([pr#15550](#), John Spray)
- doc: doc/dev: update log\_based\_pg.rst, fix some display problem ([pr#12730](#), liuchang0812)
- doc: Doc: Fixes Python Swift client commands ([issue#17746](#), [pr#12887](#), Ronak Jain)
- doc: doc/install/manual-deployment: update osd creation steps ([pr#16573](#), Sage Weil)
- doc: doc/mgr/dashboard: update dashboard docs to reflect new defaults ([pr#16241](#), Sage Weil)
- doc: doc/mon: fix ceph-authtool command in rebuild mon's sample ([pr#16503](#), huanwen ren)
- doc: doc/qa: cover config help command ([pr#16727](#), John Spray)
- doc: doc/rados.8: add offset option for put command ([pr#16155](#), Jianpeng Ma)
- doc: doc/rados: add page for health checks and update monitoring.rst ([pr#16566](#), John Spray)
- doc: doc/rados/configuration: document bluestore ([pr#16765](#), Sage Weil)
- doc: doc/radosgw/s3/cpp.rst: update usage of libs3 APIs to make the examples work ([pr#10851](#), Weibing Zhang)
- doc: doc/rados/operations/health-checks: osd section ([pr#16611](#), Sage Weil)

- doc: doc/release-notes: add Images creation timestamp note ([pr#15963](#), clove)
- doc: doc/release-notes: avoid ‘production-ready’ in describing kraken ([pr#13675](#), Sage Weil)
- doc: doc/release-notes: final kraken notes ([pr#12968](#), Sage Weil)
- doc: doc/release-notes: fix bluestore links ([pr#16787](#), Sage Weil)
- doc: doc/release-notes: fix links, formatting; add crush device class docs ([pr#16741](#), Sage Weil)
- doc: doc/release-notes: fix upmap and osd replacement links; add FIXME ([pr#16730](#), Sage Weil)
- doc: doc/release-notes: Luminous release notes typo fixes “ceph config-key ls”->“ceph config-key list” ([pr#16330](#), scienceluo)
- doc: doc/release-notes: Luminous release notes typo fixes ([pr#16338](#), Luo Kexue)
- doc: doc/release-notes: sort release note changes into the right section ([pr#16764](#), Sage Weil)
- doc: doc/release-notes: update device class cli ([pr#16851](#), xie xingguo)
- doc: doc/release-notes: update luminous notes ([pr#15851](#), Sage Weil)
- doc: doc/release-notes: update which jewel version does sortbitwise warning ([pr#15209](#), Sage Weil)
- doc: doc/releases: Update releases from Feb 2017 to July 2017 ([pr#16303](#), Bryan Stillwell)
- doc: doc/rgw: instructions for changing multisite master zone ([pr#14089](#), Casey Bodley)
- doc: doc/rgw: remove fastcgi page and sample configs ([pr#15133](#), Casey Bodley)
- doc: doc/rgw: remove Federated Configuration, clean up multisite ([issue#19504](#), [issue#18082](#), [pr#15132](#), Casey Bodley)
- doc: docs: Clarify the relationship of min\_size to EC pool recovery ([pr#14419](#), Brad Hubbard)
- doc: docs: Fix problems with example code ([pr#14007](#), Brad Hubbard)
- doc: docs: mgr dashboard ([pr#15920](#), Wido den Hollander)
- doc: [docs/quick-start]: update quick start to add a note for mgr create command for luminous+ builds ([pr#16350](#), Vasu Kulkarni)
- doc: Documentation Fixes for <http://tracker.ceph.com/issues/19879> ([issue#20057](#),

[issue#19879](#), [pr#15606](#), Sameer Tiwari)

- doc: Documentation updates for July 2017 releases ([pr#16401](#), Bryan Stillwell)
- doc: document bluestore compression settings ([pr#16747](#), Kefu Chai)
- doc: document mClock related options ([pr#16552](#), Kefu Chai)
- doc: document osd-agent-{max,low}-ops options ([pr#13648](#), Kefu Chai)
- doc: document perf histograms ([pr#15150](#), Piotr Dałek)
- doc: document “rados cleanup” in rados manpage ([issue#20894](#), [pr#16777](#), Nathan Cutler)
- doc: document repair/scrub features ([issue#15786](#), [pr#9032](#), Kefu Chai, David Zafman)
- doc: Document RGW quota cache options ([issue#18747](#), [pr#13395](#), Daniel Gryniewicz)
- doc: Document that osd\_heartbeat\_grace applies to MON and OSD ([pr#13098](#), Wido den Hollander)
- doc: document the setup of restful and dashboard plugins ([issue#20239](#), [pr#15707](#), Kefu Chai)
- doc: explain about logging levels ([pr#12920](#), liuchang0812)
- doc: fio: update README.md so only the fio ceph engine is built ([pr#15081](#), Kefu Chai)
- doc: fix a typo ([pr#13930](#), Drunkard Zhang)
- doc: fix broken link in erasure-code.rst ([issue#19972](#), [pr#15143](#), MinSheng Lin)
- doc: fix document about rados mon ([pr#12662](#), liuchang0812)
- doc: Fixed a typo in yum repo filename script ([pr#16431](#), Jeff Green)
- doc: fixes a broken hyperlink to RADOS paper in architecture ([pr#13682](#), Tahia Khan)
- doc: Fixes a typo ([pr#13985](#), Edwin F. Boza)
- doc: Fixes parameter name in rbd configuration on openstack havana/icehouse ([issue#17978](#), [pr#13403](#), Michael Eischer)
- doc: Fixes radosgw-admin ex: in swift auth section ([issue#16687](#), [pr#12646](#), SirishaGuduru)
- doc: fixes to silence sphinx-build ([pr#13997](#), Kefu Chai)
- doc: fix factual inaccuracy in doc/architecture.rst ([pr#15235](#), Nathan Cutler),

Sage Weil)

- doc: fixing an error in 12.0.3 release notes ([pr#15195](#), Abhishek Lekshmanan)
- doc: fix link for ceph-mgr cephx authorization ([pr#16246](#), Greg Farnum)
- doc: fix link that pointed to a nonexistent file ([pr#14740](#), Peter Maloney)
- doc: fix syntax on code snippets in cephfs/multimds ([pr#15499](#), John Spray)
- doc: fix the librados c api can not compile problem ([pr#9396](#), song baisen)
- doc: fix the links to <http://ceph.com/docs> ([issue#19090](#), [pr#13976](#), Kefu Chai)
- doc: Fix typo and grammar in RGW config reference ([pr#13356](#), Ruben Kerkhof)
- doc: fix typo in config.rst ([pr#16721](#), Jos Collin)
- doc: fix typos in config.rst ([pr#16681](#), Song Shun)
- doc: fix typos in radosgw-admin usage ([pr#13936](#), Enming Zhang)
- doc: freshen mgr docs ([pr#15690](#), John Spray)
- doc: hammer 0.94.10 release notes ([pr#13152](#), Nathan Cutler)
- doc: Have install put manpages in the FreeBSD correct location ([pr#13301](#), Willem Jan Withagen)
- doc: how to specify filesystem for cephfs clients ([pr#14087](#), John Spray)
- doc: improve firewalld instructions ([pr#13360](#), Ken Dreyer)
- doc: Indicate how to add multiple admin capabilities ([pr#13956](#), Chu, Hua-Rong)
- doc: instructions and guidance for multimds ([issue#19135](#), [pr#13830](#), John Spray)
- doc: instructions for provisioning OpenStack VMs ad hoc ([pr#13368](#), Nathan Cutler)
- doc: Jewel 10.2.6 release notes ([pr#13835](#), Abhishek Lekshmanan)
- doc: Jewel v10.2.8 release notes ([pr#16274](#), Nathan Cutler)
- doc: Jewel v10.2.9 release notes ([pr#16318](#), Nathan Cutler)
- doc: kernel client os-recommendations update ([pr#13369](#), John Spray, Ilya Dryomov)
- doc: kill some broken links ([pr#15203](#), liuchang0812)
- doc: kill sphinx warnings ([pr#16198](#), Kefu Chai)
- doc: luminous: doc: update rbd-mirroring documentation ([issue#20701](#), [pr#16912](#), Jason Dillaman)

- doc: Luminous release notes typo fixes ([pr#15899](#), Abhishek Lekshmanan)
- doc: mailmap: add affiliation for Zhu Shangzhong ([pr#16537](#), Zhu Shangzhong)
- doc: mailmap: add Alibaba into organization map ([pr#14900](#), James Liu)
- doc: mailmap: add Myoungwon Oh's mailmap and affiliation ([pr#15934](#), Myoungwon Oh)
- doc: mailmap for v12.0.2 ([pr#14753](#), Abhishek Lekshmanan)
- doc: mailmap: Michal Koutny affiliation ([pr#13036](#), Nathan Cutler)
- doc: mailmap, organizationmap: add affiliation for Tushar Gohad ([pr#16081](#), Tushar Gohad)
- doc: .mailmap, .organizationmap: Update Fan Yang information and affiliation ([pr#16067](#), Fan Yang)
- doc: .mailmap, .organizationmap: Update Song Weibin information and affiliation ([pr#16311](#), songweibin)
- doc: .mailmap, .organizationmap: Update ztczll affiliation ([pr#16038](#), zhanglei)
- doc: mailmap updates for v11.1.0 ([pr#12335](#), Abhishek Lekshmanan)
- doc: mailmap updates ([pr#13309](#), Loic Dachary)
- doc: mailmap: V12.0.1 credits ([pr#14479](#), M Ranga Swami Reddy)
- doc: mailmap: Willem Jan Withagen affiliation ([pr#13034](#), Willem Jan Withagen)
- doc: mailmap: ztczll affiliation ([pr#15079](#), zhanglei)
- doc: man/8/ceph-disk: fix formatting ([pr#13969](#), Kefu Chai)
- doc: mention certain conf vars should be in global ([pr#15119](#), Ali Maredia)
- doc: mention ENXIO change in the 10.2.6 release notes ([pr#13878](#), Nathan Cutler)
- doc: mention -show-mappings in crushtool manpage ([issue#19649](#), [pr#14599](#), Nathan Cutler, Loic Dachary)
- doc: mention teuthology-worker security group ([pr#14748](#), Nathan Cutler)
- doc: Merge pull request from stiwari/wip-19879 ([issue#19879](#), [pr#15609](#), Sameer Tiwari)
- doc: mgr/restful: bind to :: and update docs ([pr#16267](#), Sage Weil)
- doc: minor changes in fuse client config reference ([pr#13065](#), Barbora Ančincová)
- doc: minor change to a cloud testing paragraph ([pr#13277](#), Jan Fajerski)

- doc: minor fixes in radosgw/ ([pr#15103](#), Drunkard Zhang)
- doc: min\_size advice is not helpful ([pr#12936](#), Brad Hubbard)
- doc: misc minor fixes ([pr#13713](#), Drunkard Zhang)
- doc: Modify Configuring Cinder section ([issue#18840](#), [pr#13400](#), Shinobu Kinjo)
- doc: op queue and mclock related options ([pr#16662](#), J. Eric Ivancich)
- doc: organizationmap: add Xianxia Xiao to Kylin Cloud team ([pr#12718](#), Yunchuan Wen)
- doc: PendingReleaseNotes: “ceph -w” behavior has changed drastically ([pr#16425](#), Joao Eduardo Luis, Nathan Cutler)
- doc: PendingReleaseNotes: notes on whiteouts vs pgnls ([pr#15575](#), Sage Weil)
- doc: PendingReleaseNotes: note the fuse fstab format change ([pr#13259](#), John Spray)
- doc: PendingReleaseNotes: recent cephfs changes ([pr#14196](#), John Spray)
- doc: PendingReleaseNotes: warning about ‘osd rm ...’ and #19119 ([issue#19119](#), [pr#13731](#), Sage Weil)
- doc: peoplemap: add pdonnell alias ([pr#14352](#), Patrick Donnelly)
- doc: radosgw-admin: new ‘global quota’ commands update period config ([issue#19409](#), [pr#14252](#), Casey Bodley)
- doc: README.FreeBSD: update current status ([pr#12096](#), Willem Jan Withagen)
- doc: README.FreeBSD: Update the status ([pr#14406](#), Willem Jan Withagen)
- doc: README.md: fix build instructions inconsistent ([pr#14555](#), Yao Zongyou)
- doc: README.md: use github heading syntax to mark the headings ([pr#14591](#), Kefu Chai)
- doc: release-notes clarify about rgw encryption ([pr#14800](#), Abhishek Lekshmanan)
- doc: release notes for v10.2.7 Jewel ([pr#14295](#), Abhishek Lekshmanan)
- doc: release notes for v11.1.1 ([pr#12642](#), Abhishek Lekshmanan)
- doc: release notes for v12.0.3 (dev) ([pr#15090](#), Abhishek Lekshmanan)
- doc: releases update the luminous, hammer, jewel release dates ([pr#13584](#), Abhishek Lekshmanan)
- doc: remove deprecated subcommand in man/8/ceph.rst ([pr#14928](#), Drunkard Zhang)

- doc: remove docs on non-existant command ([pr#16616](#), Luo Kexue, Kefu Chai)
- doc: remove duplicated references ([pr#13396](#), Kefu Chai)
- doc: remove mentions about mon\_osd\_min\_down\_reports ([issue#19016](#), [pr#13558](#), Barbora Ančincová)
- doc: remove some non-existent and fix the default value according to ... ([pr#15664](#), Leo Zhang)
- doc: Remove "splitting" state ([pr#12636](#), Brad Hubbard)
- doc: reword mds deactivate docs; add optional fs\_name argument ([issue#20607](#), [pr#16471](#), Jan Fajerski)
- doc: Re-word the warnings about using git subtrees ([pr#14999](#), J. Eric Ivancich)
- doc: rgw clarify limitations when creating tenant names ([pr#16418](#), Abhishek Lekshmanan)
- doc: rgw: Clean up create subuser parameters ([pr#14335](#), hrchu)
- doc: rgw: correct get usage parameter default value ([pr#14372](#), hrchu)
- doc: rgw: Get user usage needs to specify user ([pr#14804](#), hrchu)
- doc: rgw: make a note abt system users vs normal users ([issue#18889](#), [pr#13461](#), Abhishek Lekshmanan)
- doc: rgw: note rgw\_enable\_usage\_log option in adminops guide ([pr#14803](#), hrchu)
- doc: rgw: remove mention of megabytes for quotas ([pr#14413](#), Hans van den Bogert)
- doc: rgw: Rewrite Java swift examples ([pr#14268](#), Chu, Hua-Rong)
- doc: rgw: Rewrite the key management ([pr#14384](#), hrchu)
- doc: rgw server-side encryption and barbican ([pr#13483](#), Adam Kupczyk, Casey Bodley)
- doc: script: build-doc/serve-doc fixes ([pr#14438](#), Abhishek Lekshmanan)
- doc: script: ceph-release-notes: use https instead of http ([pr#14103](#), Kefu Chai)
- docs: doc/cephfs/troubleshooting: fix broken bullet list ([pr#12894](#), Dan Mick)
- docs: doc/dev: add some info about FreeBSD ([pr#14503](#), Willem Jan Withagen)
- docs: doc/release-notes: fix ceph-deploy command ([pr#15987](#), Sage Weil)
- docs: doc/release-note: update release-note ([pr#15748](#), liuchang0812)
- docs: document "osd recovery max single start" setting ([issue#17396](#), [pr#15275](#),

Ken Dreyer)

- docs: mailmap: fix Zhao Chao affiliation ([pr#13413](#), Zhao Chao)
- docs: mailmap: Leo Zhang infomation and affiliation ([pr#15145](#), Leo Zhang)
- docs: mailmap: Liu Yang affiliation ([pr#13427](#), LiuYang)
- docs: mailmap: shiqi affiliation ([pr#14361](#), shiqi)
- docs: mailmap: update organization info ([pr#14747](#), liuchang0812)
- docs: mailmap: Weibing Zhang mailmap affiliation ([pr#15076](#), Weibing Zhang)
- docs: PendingReleaseNotes: mention forced recovery ([pr#16775](#), Piotr Dałek)
- docs: Remove contractions from the documentation ([pr#16629](#), John Wilkins)
- doc: style fix for doc/cephfs/client-config-ref.rst ([pr#14840](#), Drunkard Zhang)
- doc: tools/cephfs: fix cephfs-journal-tool -help ([pr#15614](#), John Spray)
- doc: two minor fixes ([pr#14494](#), Drunkard Zhang)
- doc: typo fixes on hyperlink/words ([pr#15144](#), Drunkard Zhang)
- doc: typo fix in s3\_compliance ([pr#12598](#), LiuYang)
- doc: typo in hit\_set\_search\_last\_n ([pr#14108](#), Sven Seeberg)
- doc: Update adminops.rst ([pr#13893](#), Chu, Hua-Rong)
- doc: update ceph(8) man page with new sub-commands ([pr#16437](#), Kefu Chai)
- doc: Update CephFS disaster recovery documentation ([pr#12370](#), Wido den Hollander)
- doc: Update disk thread section to reflect that scrubbing is no longe... ([pr#12621](#), Nick Fisk)
- doc: update intro, quick start docs ([pr#16224](#), Sage Weil)
- doc: Update keystone.rst ([pr#12717](#), Chu, Hua-Rong)
- doc: update links to point to ceph/qa instead of ceph-qa-suite ([pr#13397](#), Jan Fajerski, Nathan Cutler)
- doc: Update .organizationmap ([pr#16507](#), luokexue)
- doc: update packages mentioned by build-doc and related doc ([pr#14649](#), Yu Shengzuo)
- doc: Update sample.ceph.conf ([pr#13751](#), Saumay Agrawal)
- doc: update sample explaning "%" operator in test suites ([pr#15511](#), Kefu Chai)

- doc: Update some RGW documentation ([pr#15175](#), Jens Rosenboom)
- doc: update the pool names created by vstart.sh by default ([pr#16652](#), Zhu Shangzhong)
- doc: update the rados namespace docs ([pr#15838](#), Abhishek Lekshmanan)
- doc: update the support status of swift static website ([pr#13824](#), Jing Wenjun)
- doc: update the usage of 'ceph-deploy purge' ([pr#15080](#), Yu Shengzuo)
- doc: update to new ceph fs commands ([pr#13346](#), Patrick Donnelly)
- doc: upmap docs; various missing links for release notes ([pr#16637](#), Sage Weil)
- doc: use do\_cmake.sh instead of cmake .. ([pr#15110](#), Kefu Chai)
- doc: v12.0.0 release notes ([pr#13281](#), Abhishek Lekshmanan)
- doc: v12.0.2 (dev) release notes ([pr#14625](#), Abhishek Lekshmanan)
- doc: v12.1.0 release notes notable changes addition again ([pr#15857](#), Abhishek Lekshmanan)
- doc: various fixes ([pr#16723](#), Kefu Chai)
- doc: vstart: add -help documentation for rgw\_num ([pr#13817](#), Ali Maredia)
- doc: wip-doc-multisite ports downstream multisite document upstream ([pr#14259](#), John Wilkins)
- doc: Wip osd discussion docs ([pr#13344](#), Greg Farnum)
- filestore: os/filestore: Exclude BTRFS on FreeBSD ([pr#16171](#), Willem Jan Withagen)
- filestore: os/filestore/FileJournal: Fix typo in the comment ([pr#14493](#), Zhou Zhengping)
- filestore: os/filestore: use existing variable for same func ([pr#13742](#), Pan Liu)
- filestore: os/filestore: when print log, use \_\_func\_\_ instead of hard code function name ([pr#15261](#), mychoxin)
- filestore: os/filestore: zfs add get\_name() ([pr#15650](#), Yanhu Cao)

Fix full testing in cephtool/test.sh when used by rados suite @Jing-Scott updated, addressing @rzarzynski's change request
- librados: add log channel to rados\_monitor\_log2 callback ([pr#15926](#), Sage Weil)
- librados: add missing implementations for C service daemon API methods ([pr#16543](#), Jason Dillaman)

- librados: add override for librados ([issue#18922](#), [pr#13442](#), liuchang0812)
- librados: add override in headers ([pr#13775](#), liuchang0812)
- librados: asynchronous selfmanaged\_snap\_create/selfmanaged\_snap\_remove APIs ([issue#16180](#), [pr#12050](#), Jason Dillaman)
- librados: do not expose non-public symbols ([pr#13265](#), Kefu Chai)
- librados: fix compile errors from simplified aio completions ([pr#12849](#), xie xingguo)
- librados: fix rados\_pool\_list when buf is null ([pr#14859](#), Sage Weil)
- librados: redirect balanced reads to acting primary when targeting object isn't recovered ([issue#17968](#), [pr#15489](#), Xuehan Xu)
- librados: remove legacy object listing API, clean up newer api ([pr#13149](#), Sage Weil)
- librados: replace the var name from onack to complete ([pr#13857](#), Pan Liu)
- librados: set the flag CEPH\_OSD\_FLAG\_FULL\_TRY of Op in the right place ([pr#14193](#), Pan Liu)
- librados: use cursor for nobjects listing ([pr#13323](#), Yehuda Sadeh, Sage Weil)
- librbd: add compare and write API ([pr#14868](#), Zhengyong Wang, Jason Dillaman)
- librbd: add create timestamp metadata for image ([pr#15757](#), runsisi)
- librbd: added rbd\_flatten\_with\_progress to API ([issue#15824](#), [pr#12905](#), Ricardo Dias)
- librbd: add LIBRBD\_SUPPORTS\_WRITESAME support ([pr#16583](#), Xiubo Li)
- librbd: add override keyword in header files ([issue#19012](#), [pr#13536](#), liuchang0812)
- librbd: add SnapshotNamespace to ImageCtx ([pr#12970](#), Victor Denisov)
- librbd: add writesame API ([pr#12645](#), Mingxin Liu, Gui Hecheng)
- librbd: allow to open an image without opening the parent image ([issue#18325](#), [pr#12885](#), Ricardo Dias)
- librbd: asynchronous clone state machine ([pr#12041](#), Dongsheng Yang)
- librbd: asynchronous image removal state machine ([pr#12102](#), Dongsheng Yang, Venky Shankar)
- librbd: avoid possible recursive lock when racing acquire lock ([issue#17447](#),

- pr#12991, Jason Dillaman)
- librbd: changed the return type of ImageRequestWQ::discard() ([issue#18511](#), [pr#14032](#), Jos Collin)
  - librbd: cleanup logging code under librbd/io ([pr#14975](#), runsisi)
  - librbd: corrected resize RPC message backwards compatibility ([issue#19636](#), [pr#14615](#), Jason Dillaman)
  - librbd: create fewer empty objects during copyup ([issue#15028](#), [pr#12326](#), Douglas Fuller, Venky Shankar)
  - librbd: deferred image deletion ([issue#18481](#), [pr#13105](#), Ricardo Dias)
  - librbd: delay mirror registration when creating clones ([issue#17993](#), [pr#12839](#), Jason Dillaman)
  - librbd: discard related IO should skip op if object non-existent ([issue#19962](#), [pr#15239](#), Mykola Golub)
  - librbd: do not instantiate templates while building tests ([issue#18938](#), [pr#14891](#), Kefu Chai)
  - librbd: do not raise an error if trash list returns -ENOENT ([pr#15085](#), runsisi)
  - librbd: don't continue to remove an image w/ incompatible features ([issue#18315](#), [pr#12638](#), Dongsheng Yang)
  - librbd: eliminate compiler warnings ([pr#13729](#), Jason Dillaman)
  - librbd: fail IO request when exclusive lock cannot be obtained ([pr#15860](#), Jason Dillaman)
  - librbd: filter expected error codes from is\_exclusive\_lock\_owner ([issue#20182](#), [pr#15483](#), Jason Dillaman)
  - librbd: fix clang compilation error ([issue#19260](#), [pr#13926](#), Mykola Golub)
  - librbd: fixed initializer list ordering ([pr#13042](#), Jason Dillaman)
  - librbd: fix issues with image removal state machine ([pr#15734](#), Jason Dillaman)
  - librbd: fix rbd\_metadata\_list and rbd\_metadata\_get ([issue#19588](#), [pr#14471](#), Mykola Golub)
  - librbd: fix segfault on EOPNOTSUPP returned while fetching snapshot timestamp ([issue#18839](#), [pr#13287](#), Gui Hecheng)
  - librbd: fix valgrind errors and ensure tests detect future leaks ([pr#15415](#), Jason Dillaman)

- librbd: fix valid coverity warnings ([pr#14023](#), Jason Dillaman)
- librbd: image create validates that pool supports overwrites ([issue#19081](#), [pr#13986](#), Jason Dillaman)
- librbd: image-extent cache needs to clip out-of-bounds read buffers ([pr#13679](#), Jason Dillaman)
- librbd: Include WorkQueue.h since we use it ([issue#18862](#), [pr#13322](#), Boris Ranto)
- librbd: initialize diff parent overlap to zero ([pr#13077](#), Gu Zhongyan)
- librbd: introduce new constants for tracking max block name prefix ([issue#18653](#), [pr#13141](#), Jason Dillaman)
- librbd: is\_exclusive\_lock\_owner API should ping OSD ([issue#19287](#), [pr#14003](#), Jason Dillaman)
- librbd: managed lock refactoring ([pr#12922](#), Mykola Golub)
- librbd: metadata\_set API operation should not change global config setting ([issue#18465](#), [pr#12843](#), Mykola Golub)
- librbd: minor fixes for image trash move ([pr#14834](#), runsisi)
- librbd: new API method to force break a peer's exclusive lock ([issue#18429](#), [issue#16988](#), [issue#18327](#), [pr#12639](#), Jason Dillaman)
- librbd: Notifier::notify API improvement ([pr#14072](#), Mykola Golub)
- librbd: optimize copy-up to add hints only once to object op ([issue#19875](#), [pr#15037](#), Mykola Golub)
- librbd: pass an uint64\_t to clip\_io() as the third param ([issue#18938](#), [pr#14159](#), Kefu Chai)
- librbd: permit removal of image being bootstrapped by rbd-mirror ([issue#16555](#), [pr#12549](#), Mykola Golub)
- librbd: possible deadlock with flush if refresh in-progress ([issue#18419](#), [pr#12838](#), Jason Dillaman)
- librbd: potential read IO hang when image is flattened ([issue#19832](#), [pr#15234](#), Jason Dillaman)

- librbd: potential use of uninitialised value in ImageWatcher ([pr#14091](#), Mykola Golub)
- librbd: prevent self-blacklisting during break lock ([issue#18666](#), [pr#13110](#), Jason Dillaman)
- librbd: race initializing exclusive lock and configuring IO path ([pr#13086](#), Jason Dillaman)
- librbd: random unit test failures due to shut down race ([issue#19389](#), [pr#14166](#), Jason Dillaman)
- librbd: rbd ack cleanup ([pr#13791](#), runsis)
- librbd: reacquire lock should update lock owner client id ([issue#19929](#), [pr#15093](#), Jason Dillaman)
- librbd: reduce potential of erroneous blacklisting on image close ([issue#19970](#), [pr#15162](#), Jason Dillaman)
- librbd: refactor exclusive lock support into generic managed lock ([issue#17016](#), [pr#12846](#), Ricardo Dias, Jason Dillaman)
- librbd: relax “is parent mirrored” check when enabling mirroring for pool ([issue#19130](#), [pr#13752](#), Mykola Golub)
- librbd: remove redundant check for image id emptiness ([pr#14830](#), runsis)
- librbd: remove unnecessary dependencies of ManagedLock ([pr#12982](#), Jason Dillaman)
- librbd: remove unused rbd\_image\_options\_t ostream operator ([pr#15443](#), Mykola Golub)
- librbd: resolve static analyser warnings ([pr#12863](#), Jason Dillaman)
- librbd: scatter/gather support for the C API ([issue#13025](#), [pr#13447](#), Jason Dillaman)
- librbd: silence -Wunused-variable warning ([pr#14953](#), Kefu Chai)
- librbd: simplify image open/close semantics ([pr#13701](#), Jason Dillaman)
- librbd: support for shared locking in ManagedLock ([pr#12886](#), Ricardo Dias)
- librbd: support to list snapshot timestamp ([issue#808](#), [pr#12817](#), Pan Liu)
- librbd: Uninitialized variable used handle\_refresh() ([pr#16724](#), amitkuma)
- librbd: use ‘override’ keyword instead of ‘virtual’ ([issue#18922](#), [pr#13437](#), liuchang0812)

- librbd: warning message for mirroring pool option ([issue#18125](#), [pr#12319](#), Gaurav Kumar Garg)
- log: use one write system call per message ([pr#11955](#), Patrick Donnelly)
- mds: add authority check for delay dirfrag split ([issue#18487](#), [pr#12994](#), "Yan, Zheng")
- mds: add override in headers ([pr#13691](#), liuchang0812)
- mds: add override in mds subsystem ([issue#18922](#), [pr#13438](#), liuchang0812)
- mds: add perf counters for file system operations ([pr#14938](#), Michael Sevilla)
- mds: automate MDS object count tracking ([pr#13591](#), Patrick Donnelly)
- mds: bump client\_reply debug to match client\_req ([pr#14036](#), Patrick Donnelly)
- mds: change\_attr++ and set ctime for set\_vxattr ([issue#19583](#), [pr#14726](#), Patrick Donnelly)
- mds: change the type of data\_pools ([pr#15278](#), Vicente Cheng)
- mds: check export pin during replay ([issue#20039](#), [pr#15205](#), Patrick Donnelly)
- mds: check for errors decoding backtraces ([issue#18311](#), [pr#12588](#), John Spray)
- mds: Client syncfs is slow (waits for next MDS tick) ([issue#20129](#), [pr#15544](#), dongdong tao)
- mds: don't assert on read errors in RecoveryQueue ([issue#19282](#), [pr#14017](#), John Spray)
- mds: don't modify inode that is not projected ([issue#16768](#), [pr#13052](#), "Yan, Zheng")
- mds: drop partial entry and adjust write\_pos when opening PurgeQueue ([issue#19450](#), [pr#14447](#), "Yan, Zheng")
- mds: explicitly output error msg for dump cache asok command ([pr#15592](#), Zhi Zhang)
- mds: extend 'p' auth cap to cover all vxattr stuff ([issue#19075](#), [pr#13628](#), John Spray)
- mds: finish clientreplay requests before requesting active state ([issue#18461](#), [pr#12852](#), Yan, Zheng)
- mds: fix bad iterator dereference reported by coverity ([issue#18830](#), [pr#13272](#), John Spray)
- mds: fix CDir::merge() for mds\_debug\_auth\_pins ([issue#19946](#), [pr#15130](#), "Yan,

Zheng")

- mds: fix client ID truncation ([pr#15258](#), Henry Chang)
- mds: fix handling very fast delete ops ([issue#19245](#), [pr#13899](#), John Spray)
- mds: fix hangs involving re-entrant calls to journaler ([issue#20165](#), [pr#15430](#), John Spray)
- mds: fix incorrect assertion in Server::\_dir\_is\_nonempty() ([issue#18578](#), [pr#12973](#), Yan, Zheng)
- mds: fix IO error handling in SessionMap ([pr#13464](#), John Spray)
- mds: fix mantle script to not fail for last rank ([issue#19589](#), [pr#14704](#), Patrick Donnelly)
- mds: fix mgrc shutdown ([issue#19566](#), [pr#14505](#), John Spray)
- mds: fix null pointer dereference in Locker::handle\_client\_caps ([issue#18306](#), [pr#12808](#), Yan, Zheng)
- mds: fix stray creation/removal notification ([issue#19630](#), [pr#14554](#), "Yan, Zheng")
- mds: fix use-after-free in Locker::file\_update\_finish() ([issue#19828](#), [pr#14991](#), "Yan, Zheng")
- mds: ignore ENOENT on writing backtrace ([issue#19401](#), [pr#14207](#), John Spray)
- mds: ignore fs full check for CEPH\_MDS\_OP\_SETFILELOCK ([issue#18953](#), [pr#13455](#), "Yan, Zheng")
- mds: improvements for stray reintegration ([pr#15548](#), "Yan, Zheng")
- mds: include advisory path field in damage ([issue#18509](#), [pr#14104](#), John Spray)
- mds: issue new caps when sending reply to client ([issue#19635](#), [pr#14743](#), "Yan, Zheng")
- mds: limit client writable range increment ([issue#19955](#), [pr#15131](#), "Yan, Zheng")
- mds: make C\_MDSInternalNoop::complete() delete 'this' ([issue#19501](#), [pr#14347](#), "Yan, Zheng")
- mds: mds perf item 'l\_mdl\_expos' always behind journaler ([pr#15621](#), redickwang)
- mds: miscellaneous fixes ([issue#18646](#), [pr#12974](#), Yan, Zheng, "Yan, Zheng")
- mds: miscellaneous multimds fixes ([issue#19022](#), [pr#13698](#), "Yan, Zheng")
- mds: miscellaneous multimds fixes part2 ([pr#15125](#), "Yan, Zheng")

- mds: miscellaneous multimds fixes ([pr#14550](#), "Yan, Zheng")
- mds: misc multimds fixes ([issue#18717](#), [issue#18754](#), [pr#13227](#), "Yan, Zheng")
- mds: misc multimds fixes part2 ([pr#12794](#), Yan, Zheng)
- mds: misc multimds fixes ([pr#12274](#), Yan, Zheng)
- mds,mon: Clean issues detected by cppcheck ([pr#13199](#), Ilya Shipitsin)
- mds: multimds flock fixes ([pr#15440](#), "Yan, Zheng")
- mds: Pass empty string to clear mantle balancer ([issue#20076](#), [pr#15282](#), Zhi Zhang)
- mds: pretty json from tell commands ([pr#14105](#), John Spray)
- mds: print rank as int ([issue#19201](#), [pr#13816](#), Patrick Donnelly)
- mds: propagate error encountered during opening inode by number ([issue#18179](#), [pr#12749](#), Yan, Zheng)
- mds: properly create aux subtrees for pinned directory ([issue#20083](#), [pr#15300](#), "Yan, Zheng")
- mds: relocate PTRWAITER put near get ([pr#14921](#), Patrick Donnelly)
- mds: remove boost::pool usage and use tcmalloc directly ([issue#18425](#), [pr#12792](#), Zhi Zhang)
- mds: remove legacy "mds tell" command ([issue#19288](#), [pr#14015](#), John Spray)
- mds: remove "mds log" config option ([issue#18816](#), [pr#14652](#), John Spray)
- mds: remove some redundant object counters ([pr#13704](#), Patrick Donnelly)
- mds: replace C\_VoidFn in MDSDaemon with lambdas ([pr#13465](#), John Spray)
- mds: Return error message instead of asserting ([pr#14469](#), Brad Hubbard)
- mds: save projected path into inode\_t::stray\_prior\_path ([issue#20340](#), [pr#15800](#), "Yan, Zheng")
- mds: set ceph-mds name uncond for external tools ([issue#19291](#), [pr#14021](#), Patrick Donnelly)
- mds: shut down finisher before objecter ([issue#19204](#), [pr#13859](#), John Spray)
- mds: skip fragment space check for replayed request ([issue#18660](#), [pr#13095](#), "Yan, Zheng")
- mds: support export pinning on directories ([issue#17834](#), [pr#14598](#), "Yan, Zheng", Patrick Donnelly)

- mds: try to avoid false positive heartbeat timeouts ([issue#19118](#), [pr#13807](#), John Spray)
- mds: use debug\_mds for most subsys ([issue#19734](#), [pr#15052](#), Patrick Donnelly)
- mds: use same inode count in health check as in trim ([issue#19395](#), [pr#14197](#), John Spray)
- mds: warn if insufficient standbys exist ([issue#17604](#), [pr#12074](#), Patrick Donnelly)
- mgr: add a get\_version to the python interface ([pr#13669](#), John Spray)
- mgr: add machinery for python modules to send MCommands to daemons ([pr#14920](#), John Spray)
- mgr: add mgr allow \* to client.admin ([pr#14864](#), huanwen ren)
- mgr: add override in headers ([pr#13772](#), liuchang0812)
- mgr: add override in mgr subsystem ([issue#18922](#), [pr#13436](#), liuchang0812)
- mgr: add per-DaemonState lock ([pr#16432](#), Sage Weil)
- mgr: always free allocated MgrPyModule ([issue#19590](#), [pr#14507](#), Kefu Chai)
- mgr: ceph-create-keys: update client.admin if it already exists ([issue#19940](#), [pr#15112](#), John Spray)
- mgr: ceph: introduce “tell x help” subcommand ([issue#19885](#), [pr#15111](#), liuchang0812)
- mgr: ceph-mgr: Implement new pecan-based rest api ([pr#14457](#), Boris Ranto)
- mgr: ceph-mgr: rotate logs on sighup ([issue#19568](#), [pr#14437](#), Dan van der Ster)
- mgr: clean up daemon start process ([issue#20383](#), [pr#16020](#), John Spray)
- mgr: clean up fsstatus module ([pr#15925](#), John Spray)
- mgr: cleanup, stop clients sending in perf counters ([pr#15578](#), John Spray)
- mgr: cluster log message on plugin load error ([pr#15927](#), John Spray)
- mgr: dashboard code cleanup ([pr#15577](#), John Spray)
- mgr: dashboard GUI module ([pr#14946](#), John Spray, Dan Mick)
- mgr: dashboard improvements ([pr#16043](#), John Spray)
- mgr: do shutdown using finisher so we can do it in the right order ([issue#19743](#), [pr#14835](#), Kefu Chai)

- mgr: do the shutdown in the right order ([issue#19813](#), [pr#14952](#), Kefu Chai)
- mgr: drop repeated log info. and unnecessary write permission ([pr#15896](#), Yan Jun)
- mgr: enable ceph\_send\_command() to send pg command ([pr#15865](#), Kefu Chai)
- mgr: fix bugs in init, beacons ([issue#19516](#), [issue#19502](#), [pr#14374](#), Sage Weil)
- mgr: fix crash on missing 'ceph\_version' in daemon metadata (fixes #18764) ([issue#18764](#), [pr#14129](#), Tim Serong)
- mgr: fix crash on set\_config from python module with insufficient caps ([issue#19629](#), [pr#14706](#), Tim Serong)
- mgr: fix lock cycle ([pr#16508](#), Sage Weil)
- mgr: fix metadata handling from old MDS daemons ([pr#14161](#), John Spray)
- mgr: fix MgrStandby eating messages ([pr#15716](#), John Spray)
- mgr: fix python module teardown & add tests ([issue#19407](#), [issue#19412](#), [issue#19258](#), [pr#14232](#), John Spray)
- mgr: fix session leak ([issue#19591](#), [pr#14720](#), Sage Weil)
- mgr: fix several init/re-init bugs ([issue#19491](#), [pr#14328](#), Sage Weil)
- mgr: handle "module.set\_config(.., None)" correctly ([pr#16749](#), Kefu Chai)
- mgr: increase debug level for ticks 0 -> 10 ([pr#16301](#), Dan Mick)
- mgr: load modules in separate python sub-interpreters ([pr#14971](#), Tim Serong)
- mgr: luminous: mgr: add missing call to pick\_addresses ([issue#20955](#), [issue#21049](#), [pr#17173](#), John Spray)
- mgr: Make stats period configurable ([issue#17449](#), [pr#12732](#), liuchang0812)
- mgr: Mark session connections down on shutdown ([issue#19900](#), [pr#15192](#), Brad Hubbard)
- mgr: mgr/ClusterState: do not mangle PGMap outside of Incremental ([issue#20208](#), [pr#16262](#), Sage Weil)
- mgr: mgr/DaemonServer.cc: log daemon type string as well as id ([pr#15560](#), Dan Mick)
- mgr: mgr/dashboard: add OSD list view ([pr#16373](#), John Spray)
- mgr: mgr/dashboard: fix type error in get\_rate function ([issue#20276](#), [pr#15668](#), liuchang0812)
- mgr: mgr/dashboard: load log lines on startup, split out audit log ([pr#15709](#),

John Spray)

- mgr: mgr/MgrClient: fix reconnect event leak ([issue#19580](#), [pr#14431](#), Sage Weil)
- mgr: mgr/MgrStandby: prevent use-after-free on just-shut-down Mgr ([issue#19595](#), [pr#15297](#), Sage Weil)
- mgr: mgr/MgrStandby: respawn when deactivated ([issue#19595](#), [issue#19549](#), [pr#15557](#), Sage Weil)
- mgr: mgr\_module interface to report health alerts ([pr#16487](#), Sage Weil)
- mgr: mgr,osd: ceph-mgr -help, unify usage text of other daemons ([pr#15176](#), Tim Serong)
- mgr: mgr/PyState: shut up about get\_config on nonexistent keys ([pr#16641](#), Sage Weil)
- mgr: mgr/status: row has incorrect number of values ([issue#20750](#), [pr#16529](#), liuchang0812)
- mgr: Misc. bug fixes ([issue#18994](#), [pr#14883](#), John Spray)
- mgr: mkdir bootstrap-mgr ([pr#14824](#), huanwen ren)
- mgr: mon/mgr: add detail error infomation ([pr#16048](#), Yan Jun)
- mgr,mon: mgr,mon: debug init and mgrdigest subscriptions ([issue#20633](#), [pr#16351](#), Sage Weil)
- mgr: mon/MgrMonitor: fix standby addition to mgrmap ([issue#20647](#), [pr#16397](#), Sage Weil)
- mgr,mon: mon/AuthMonitor: generate bootstrap-mgr key on upgrade ([issue#20666](#), [pr#16395](#), Joao Eduardo Luis)
- mgr,mon: mon,mgr: extricate PGmap from monitor ([issue#20067](#), [issue#20174](#), [issue#20050](#), [pr#15073](#), Kefu Chai, Sage Weil, Greg Farnum)
- mgr,mon: mon/MgrMonitor: add ‘mgr dump [epoch]’ command ([pr#15158](#), Sage Weil)
- mgr,mon: mon/MgrMonitor: only propose if we updated ([pr#14645](#), Sage Weil)
- mgr,mon: mon/MgrMonitor: reset mgrdigest timer with new subscription ([issue#20633](#), [pr#16582](#), Sage Weil)
- mgr,mon: mon,mgr: move reweight-by-\* to mgr ([pr#14404](#), Kefu Chai)
- mgr,mon: mon,mgr: print pgmap reports to debug (not cluster) log ([pr#15740](#), Sage Weil)
- mgr,mon: mon,mgr: trim osdmap without the help of pgmap ([pr#14504](#), Kefu Chai)

- mgr: move ‘osd perf’ and ‘osd blocked-by’ to mgr ([pr#14303](#), Sage Weil)
- mgr: move “osd pool stats” to mgr ([pr#14365](#), Kefu Chai)
- mgr: optimization some judgment and adjust the debug remove value in register\_new\_pgs ([pr#14046](#), song baisen)
- mgr: optimize DaemonStateIndex::cull() a little bit ([pr#14967](#), Kefu Chai)
- mgr: pass through cluster log to plugins ([pr#13690](#), John Spray)
- mgr: perf schema fns/change notification and Prometheus plugin ([pr#16406](#), Dan Mick)
- mgr: print a more helpful error message for when users lack mgr ceph caps ([issue#20296](#), [pr#15697](#), Greg Farnum)
- mgr,pybind: luminous: mgr/dashboard: fix duplicate images listed on iSCSI status page ([issue#21017](#), [pr#17282](#), Jason Dillaman)
- mgr: pybind/mgr/dashboard: bind to :: by default ([pr#16223](#), Sage Weil)
- mgr: pybind/mgr/dashboard: monkeypatch os.exit to stop cherrypy from taking down mgr ([issue#20216](#), [pr#15588](#), Sage Weil)
- mgr: pybind/mgr: Delete rest module ([pr#15429](#), John Spray)
- mgr: pybind/mgr/rest: completely terminate cherrypy in shutdown ([pr#14995](#), Tim Serong)
- mgr: pybind/mgr/rest: don’t set timezone to Chicago ([pr#14184](#), Tim Serong)
- mgr: pybind/mgr/restful: improve cert handling; work with vstart ([pr#15405](#), Sage Weil)
- mgr: pybind/mgr/zabbix: fix health in non-compat mode ([issue#20767](#), [pr#16580](#), Sage Weil)
- mgr,pybind,rbd: mgr/dashboard: show rbd image features ([pr#16468](#), Yanhu Cao)
- mgr: raise python exception on failure in send\_command() ([pr#15704](#), Kefu Chai)
- mgr,rbd: mgr/dashboard: RBD iSCSI daemon status page ([pr#16547](#), Jason Dillaman)
- mgr,rbd: mgr/dashboard: rbd mirroring status page ([pr#16360](#), Jason Dillaman)
- mgr,rbd: pybind/mgr/dashboard: initial block integration ([pr#15521](#), Jason Dillaman)
- mgr: redirect python stdout,stderr to ceph log ([pr#14189](#), Kefu Chai, Tim Serong, Dan Mick)

- mgr: release allocated PyString ([pr#14716](#), Kefu Chai)
- mgr: remove default cert; disable both restful and dashboard by default ([pr#15601](#), Boris Ranto, Sage Weil)
- mgr: remove non-existent MDS daemons from FSMap ([issue#17453](#), [pr#14937](#), Spandan Kumar Sahu)
- mgr: remove unused function declarations ([pr#14366](#), Wei Jin)
- mgr: rm nonused main function ([pr#14313](#), Wei Jin)
- mgr: shutdown py\_modules in Mgr::shutdown() ([issue#19258](#), [pr#14078](#), Kefu Chai)
- mgr,tests: qa/suites: move mgr tests into rados suite ([pr#14687](#), John Spray)
- mgr,tests: qa/upgrade/jewel-x/point-to-point: add a mgr during final upgrade ([pr#15637](#), Sage Weil)
- mgr: use unique\_ptr for MgrStandby::active\_mgr ([pr#13667](#), John Spray)
- mgr: various cleanups ([pr#14802](#), Kefu Chai)
- mgr: vstart.sh: fix mgr vs restful command startup race ([pr#16564](#), Sage Weil)
- mgr: Zabbix monitoring module ([pr#16019](#), Wido den Hollander)
- misc: fix code typos in header files ([pr#12716](#), Xianxia Xiao)
- misc: kill clang warnings ([pr#14549](#), Kefu Chai)
- misc: Warning Elimination ([pr#14439](#), Adam C. Emerson)
- mon: add crush type down health warnings ([pr#14914](#), Neha Ojha)
- mon: added bootstrap-rbd auth profile ([pr#16633](#), Jason Dillaman)
- mon: add force-create-pg back ([issue#20605](#), [pr#16353](#), Kefu Chai)
- mon: add mgr metdata commands, and overall 'versions' command for all daemon versions ([pr#16460](#), Sage Weil)
- mon: add mon\_debug\_no\_require\_luminous ([pr#14490](#), Sage Weil)
- mon: Add override for FsNewHandler::handle() ([pr#15331](#), yonghengdexin735)
- mon: add override in headers ([pr#13693](#), liuchang0812)
- mon: add override in mon subsystem ([issue#18922](#), [pr#13440](#), liuchang0812)
- mon: add support public\_bind\_addr option ([pr#16189](#), Bassam Tabbara)
- mon: add warn info for osds were removed from osdmap but still kept in crushmap

- ([pr#12273](#), song baisen)
- mon: a few health fixes ([pr#16415](#), xie xingguo)
- mon: a few more upmap (and other) fixes ([pr#16239](#), xie xingguo)
- mon: avoid segfault in wait\_auth\_rotating ([issue#19566](#), [pr#14430](#), John Spray)
- mon: avoid start election twice when quorum enter ([pr#10150](#), song baisen)
- mon: check is\_shutdown() in timer callbacks ([issue#19825](#), [pr#14919](#), Kefu Chai)
- mon: clean up in ceph\_mon.cc ([pr#14102](#), huanwen ren)
- mon: clean up some osdmon/pgmon interactions ([pr#12403](#), Sage Weil)
- mon: cleanups ([pr#15272](#), Kefu Chai)
- mon: collect mon metdata as part of the election ([issue#20434](#), [pr#16148](#), Sage Weil)
- mon: common/config\_opts.h: kill mon\_pg\_create\_interval ([pr#13800](#), xie xingguo)
- mon: 'config-key put' -> 'config-key set' ([pr#16569](#), Sage Weil)
- mon: crush straw\_calc\_version value is 0 or 1 not 0 to 2 ([pr#13554](#), song baisen)
- mon: debug session feature tracking ([issue#20475](#), [pr#16128](#), Sage Weil)
- mon: delete unused config opts of mon\_sync\_fs\_threshold ([pr#15676](#), linbing)
- mon: delete useless function definition ([pr#15188](#), shiqi)
- mon: detect existing fs and duplicate name earlier ([issue#18964](#), [pr#13471](#), Patrick Donnelly)
- mon: DIVIDE\_BY\_ZERO in PGMapDigest::dump\_pool\_stats\_full() ([pr#15622](#), Jos Collin)
- mon: Division by zero in PGMapDigest::dump\_pool\_stats\_full() ([pr#15901](#), Jos Collin)
- mon: do crushtool test with fork and timeout, but w/o exec of crushtool ([issue#19964](#), [pr#16025](#), Sage Weil)
- mon: do not dereference empty mgr\_commands ([pr#16501](#), Sage Weil)
- mon: do not prime\_pg\_temp creating pgs; clean up pg create conditions ([issue#19826](#), [pr#14913](#), Sage Weil)
- mon: don't call propose\_pending in prepare\_update() ([issue#19738](#), [pr#14711](#), John Spray)
- mon: don't kill MDSs unless some beacons are getting through ([issue#19706](#),

pr#15308, John Spray)

- mon: don't prefix mgr summary with epoch number (pr#15512, John Spray)
- mon: don't set last\_osd\_report when the pg stats msg is ignored (pr#12975, Zhiqiang Wang)
- mon: drop useless assignment statements (pr#13958, wangzhengyong)
- mon: emit cluster log messages on MDS health changes (issue#19551, pr#14398, John Spray)
- mon: enable luminous monmap feature on full quorum (pr#13379, Joao Eduardo Luis)
- mon: extensible output format for health checks (pr#16701, John Spray)
- mon: Filter log last output by severity and channel (pr#15924, John Spray)
- mon: fix accesing pending\_fsmap from peon (issue#20040, pr#15213, John Spray)
- mon: fix a few bugs with the osd health reporting (pr#15179, Sage Weil)
- mon: fix a few nits (pr#12670, Sage Weil)
- mon: Fix deep\_age copy paste error (pr#16434, Brad Hubbard)
- mon: Fixed typo in function comment blocks and in other comments (pr#15304, linbing)
- mon: Fixed typo in @post of \_active() (pr#15191, Linbing)
- mon: fix force\_pg\_create pg stuck in creating bug (issue#18298, pr#12539, Sage Weil)
- mon: fix hang on deprecated/removed 'pg set\_full\_ratio' commands (issue#20600, pr#16300, Sage Weil)
- mon: fix hiding mdsmonitor informative strings (issue#16709, pr#13904, John Spray)
- mon: fix kvstore type in mon compact command (pr#15954, liuchang0812)
- mon: fix legacy health checks in 'ceph status' during upgrade; fix jewel-x upgrade combo (pr#17176, Sage Weil)
- mon: fix mon\_keyvaluedb application (pr#15059, Sage Weil)
- mon: Fix output text and doc (pr#16367, Yan Jun)
- mon: fix prime\_pg\_temp overrun (issue#19874, pr#14979, Sage Weil)
- mon: Fix status output warning for mon\_warn\_osd\_usage\_min\_max\_delta (issue#20544, pr#16220, David Zafman)

- mon: fix synchronise pgmap with others ([pr#14418](#), song baisen, z09440)
- mon: fix wrongly delete routed pgstats op ([issue#18458](#), [pr#12784](#), Mingxin Liu)
- mon: fix wrong mon-num counting logic of ‘ceph features’ command ([pr#17172](#), xie xingguo)
- mon: handle cases where store->get() may return error ([issue#19601](#), [pr#14678](#), Jos Collin)
- mon: include device class in tree view; hide shadow hierarchy ([pr#16016](#), Sage Weil)
- mon: Incorrect expression in PGMap::get\_health() ([pr#15648](#), Jos Collin)
- mon: in output of “ceph osd df tree”, display “-”, not “0”, for pg amount of a bucket ([pr#13015](#), Chuanhong Hong)
- mon: it’s no need to get pg action\_primary osd twice in pg scrub ([pr#15313](#), linbing)
- mon: ‘\* list’ -> ‘\* ls’ ([pr#16423](#), Sage Weil)
- mon: load mgr commands at runtime ([pr#16028](#), John Spray, Sage Weil)
- mon: logclient: use the seq id of the 1st log entry when resetting session ([issue#19427](#), [pr#14927](#), Kefu Chai)
- mon: Log errors at startup ([issue#14088](#), [pr#15723](#), Brad Hubbard)
- mon: luminous: mon/MonCommands: fix copy-and-paste error ([pr#17274](#), xie xingguo)
- mon: maintain the “cluster” PerfCounters when using ceph-mgr ([issue#20562](#), [pr#16249](#), Greg Farnum)
- mon: mark osd create as deprecated ([pr#15641](#), Joao Eduardo Luis)
- mon: mon,crush: create crush rules using device classes for replicated and ec pools via cli ([pr#16027](#), Sage Weil)
- mon: mon/HealthMonitor: avoid sending unnecessary MMonHealthChecks to leader ([pr#16478](#), xie xingguo)
- mon: mon/HealthMonitor: trigger a proposal if stat updated ([pr#16477](#), Kefu Chai)
- mon: mon/LogMonitor: don’t read list’s end() for log last ([pr#16376](#), Joao Eduardo Luis)
- mon: mon/MDSMonitor: close object section of formatter ([pr#16516](#), Chang Liu)
- mon: mon/MDSMonitor: remove create\_new\_fs from header ([pr#14019](#), Henrik Korkuc)

- mon: mon/MgrMonitor: only induce mgr epoch shortly after mkfs ([pr#16356](#), Sage Weil)
- mon: mon/MgrMonitor: send digests only if `is_active()` ([pr#15109](#), Kefu Chai)
- mon: mon/MgrStatMonitor: do not crash on luminous dev version upgrades ([pr#16287](#), Sage Weil)
- mon: mon/MonClient: cancel pending commands on shutdown ([issue#20051](#), [pr#15227](#), Kefu Chai, Sage Weil)
- mon: mon/MonClient: make `get_mon_log_message()` atomic ([issue#19427](#), [pr#14422](#), Kefu Chai)
- mon: mon/MonClient: random all ranks then pick `first_n` ([pr#13479](#), Mingxin Liu)
- mon: mon/Monitor.h: add const to member function ([pr#10412](#), Michal Jarzabek)
- mon: mon/Monitor: recreate mon session if features changed ([issue#20433](#), [pr#16230](#), Joao Eduardo Luis)
- mon: {mon,osd,mds} {versions,count-metadata} ([pr#15436](#), Sage Weil)
- mon: mon/OSDMonitor: a couple of upmap and other fixes ([pr#15917](#), xie xingguo)
- mon: mon/OSDMonitor: check `get()`'s return value instead of bl's length ([pr#14805](#), Kefu Chai)
- mon: mon/OSDMonitor: check `last_osd_report` only when the whole cluster is lu... ([pr#14294](#), Kefu Chai)
- mon: mon/OSDMonitor: Clean up: delete extra S signature for plural ([pr#14174](#), Shinobu Kinjo)
- mon: mon/OSDMonitor: cleanup `pending_created_pgs` after done with it ([pr#14898](#), Kefu Chai)
- mon: mon/OSDMonitor: do not alter the “created” epoch of a pg ([issue#19787](#), [pr#14849](#), Kefu Chai)
- mon: mon/OSDMonitor: ensure UP is not set for newly-created OSDs ([issue#20751](#), [pr#16534](#), Sage Weil)
- mon: mon/OSDMonitor: fix dividing by zero in `OSDUtilizationDumper` ([pr#13531](#), Mingxin Liu)
- mon: mon/OSDMonitor: fix output func name in `can_mark_out` ([pr#14758](#), xie xingguo)
- mon: mon/OSDMonitor: fix process osd failure ([pr#12938](#), Mingxin Liu)
- mon: mon/OSDMonitor: guard ‘osd crush set-device-class’ ([pr#16217](#), Sage Weil)

- mon: mon/OSDMonitor: increase last\_epoch\_clean's lower bound if possible ([pr#14855](#), Kefu Chai)
- mon: mon/OSDMonitor: issue pool application related warning ([pr#16520](#), xie xingguo)
- mon: mon/OSDMonitor: “osd crush class rename” support ([pr#15875](#), xie xingguo)
- mon: mon/OSDMonitor: remove trivial PGMap dependency for ‘osd primary-temp’ command ([pr#13616](#), Sage Weil)
- mon: mon/OSDMonitor: remove zeroed new\_state updates ([issue#20751](#), [pr#16518](#), Sage Weil)
- mon: mon/OSDMonitor: sanity check osd before performing ‘osd purge’ ([pr#16838](#), xie xingguo)
- mon: mon/OSDMonitor: some cleanup for reweight-by-pg ([pr#13462](#), Haodong Tang)
- mon: mon/OSDMonitor: spinlock -> std::mutex ([pr#14269](#), Sage Weil)
- mon: mon/OSDMonitor: tolerate upgrade from post-kraken dev cluster ([pr#14442](#), Sage Weil)
- mon: mon/OSDMonitor: transit creating\_pgs from pgmap when upgrading ([issue#19584](#), [pr#14551](#), Kefu Chai)
- mon: mon/OSDMonitor: two pool opts related fix ([pr#15968](#), xie xingguo)
- mon: mon/OSDMonitor: update creating epoch if target osd changed ([issue#19515](#), [pr#14386](#), Kefu Chai)
- mon: mon/OSDMonitor: update creating\_pgs using pending\_creatings ([issue#19814](#), [pr#14897](#), Kefu Chai)
- mon: mon/OSDMonitor: update pg\_creatings even the new acting set is empty ([issue#19744](#), [pr#14730](#), Kefu Chai)
- mon: mon/PaxosService: use \_\_func\_\_ instead of hard code function name ([pr#15863](#), Yanhu Cao)
- mon: mon/PGMap: add up\_primary pg number field for pg-dump cmd ([pr#13451](#), xie xingguo)
- mon: mon/PGMap.cc: fix “osd\_epochs” section of dump\_basic ([pr#14996](#), xie xingguo)
- mon: mon/PGMap: make si units more readable in PGMap summary ([pr#14185](#), liuhong)
- mon: mon/PGMap: remove skewed utilization warning ([issue#20730](#), [pr#16461](#), Sage Weil)
- mon: mon/PGMap: show %used in formatted output ([issue#20123](#), [pr#15387](#), Joao

Eduardo Luis)

- mon: mon/PGMonitor: clean up min/max span warning ([pr#14611](#), Sage Weil)
- mon: mon/PGMonitor: fix description for ceph pg ls ([pr#12807](#), runsisi)
- mon: mon/PGMonitor: rm nonused function ([pr#14033](#), Wei Jin)
- mon: move ‘pg map’ to OSDMonitor ([pr#14559](#), Sage Weil)
- mon: no delay for single message MSG\_ALIVE and MSG\_PGTEMP ([pr#12107](#), yaoning)
- mon: optracker’s initiated\_at timestamp should not be NULL ([pr#12826](#), Mingxin Liu)
- mon: osd crush set crushmap need sanity check ([issue#19302](#), [pr#14029](#), Loic Dachary)
- mon: OSDMonitor add check only concern our self cluster command ([pr#10309](#), song baisen)
- mon/OSDMonitor: add plain output for “crush class ls-osd” command ([pr#17230](#), xie xingguo)
- mon/OSDMonitor: check creating\_pgs.last\_scan\_epoch instead when sending creates ([issue#20785](#), [pr#17257](#), Kefu Chai)
- mon: OSDMonitor: check mon\_max\_pool\_pg\_num when set pool pg\_num ([pr#16511](#), chenhg)
- mon/OSDMonitor: do not send\_pgCreates with stale info ([issue#20785](#), [pr#17191](#), Kefu Chai)
- mon/OSDMonitor: fix improper input/testing range of crush somke testing ([pr#17232](#), xie xingguo)
- mon: osd/PGMonitor: always update pgmap with latest osdmap ([issue#19398](#), [pr#14777](#), Kefu Chai)
- mon/pgmap: add objects prefix for unfound type ([issue#21127](#), [pr#17264](#), huanwen ren, Sage Weil)
- mon/PGMap: fix “0 stuck requests are blocked > 4096 sec” warn ([pr#17215](#), xie xingguo)
- mon: PGMonitor add check only concern our self cluster command ([pr#9976](#), song baisen)
- mon: post-jewel cleanups ([pr#13150](#), Kefu Chai)
- mon: prime pg\_temp and a few health warning fixes ([pr#16530](#), xie xingguo)

- mon: refactor MDSMonitor command handling ([pr#13581](#), John Spray)
- mon: Removed unnecessary function declaration in MDSMonitor.h ([pr#15374](#), yonghengdexin735)
- mon: remove the redundant judgement in paxosservice is\_writeable function ([pr#10240](#), song baisen)
- mon: remove unnecessary function declaration ([pr#13762](#), liuchang0812)
- mon: replace osds with osd destroy and osd new ([pr#14074](#), Joao Eduardo Luis, Sage Weil)
- mon: restructure prime\_pg\_temp around a full pg mapping calculated on multiple CPUs ([pr#13207](#), Sage Weil)
- mon: revamp health check/warning system ([pr#15643](#), John Spray, Sage Weil)
- mon: revise “ceph status” output ([pr#15396](#), John Spray)
- mon: show class in ‘osd crush tree’ output; sort output ([pr#16740](#), Sage Weil)
- mon: show destroyed status in tree view; do not auto-out destroyed osds ([pr#16446](#), xie xingguo)
- mon: show inactive % in ceph status ([pr#14810](#), Sage Weil)
- mon: show io status quickly if no update in a long period ([pr#14176](#), Mingxin Liu)
- mon: show the leader info on mon stat command ([pr#14178](#), song baisen)
- mon: skip crush smoke test when running under valgrind ([issue#20602](#), [pr#16346](#), Sage Weil)
- mon: smooth io/recovery stats over longer period ([pr#13249](#), Sage Weil)
- mon: stop issuing not-[deep]-scrubbed warnings if disabled ([pr#16465](#), xie xingguo)
- mon: support pool application metadata key/values ([pr#15763](#), Jason Dillaman)
- mon,tests: qa/suites: add test exercising workunits/mon/auth\_caps.sh ([pr#15754](#), Kefu Chai)
- mon,tests: test: Initialize pointer msg in MonClientHelper ([pr#16784](#), amitkuma)
- mon: Tidy up removal of debug mon features ([pr#14467](#), Brad Hubbard)
- mon: track features from connect clients, and use it to gate set-require-min-compat-client ([pr#15371](#), Sage Weil)
- mon: trim the creating\_pgs after updating it with pgmap ([issue#20067](#), [pr#15318](#),

Kefu Chai)

- mon: update mgrmap when active goes offline ([issue#19407](#), [pr#14220](#), John Spray)
- mon: Update OSDMon.cc comments ([pr#13750](#), Saumay Agrawal)
- mon: warn about using osd new instead of osd create ([pr#17302](#), Neha Ojha)
- msg: allow different ms type for cluster network and public network ([pr#12023](#), Haomai Wang)
- msg: always set header.version in encode\_payload() ([issue#19939](#), [pr#16421](#), Kefu Chai)
- msg: client bind ([pr#12901](#), Zengran Zhang, Haomai Wang)
- msg: do not enable client-side binding by default ([issue#20049](#), [pr#15392](#), Jason Dillaman)
- msg: don't set msgr addr when disabling client bind ([pr#15243](#), Haomai Wang)
- msg: end parameter in entity\_addr\_t::parse is optional ([pr#13650](#), Mykola Golub)
- msg: Fix calls to Messenger::create with new parameter ([pr#13329](#), Sarit Zubakov)
- msg: Increase loglevels on some messages ([pr#14707](#), Willem Jan Withagen)
- msg: Initialize member variables in Infiniband ([pr#16781](#), amitkuma)
- msg: Initializing uninitialized members MMonGetVersion ([pr#16811](#), amitkuma)
- msg: Initializing uninitialized members MMonGetVersionReply ([pr#16813](#), amitkuma)
- msg: Initializing uninitialized members MMonPaxos ([pr#16814](#), amitkuma)
- msg: Initializing uninitialized members MMonProbe ([pr#16815](#), amitkuma)
- msg: Initializing uninitialized members module messages ([pr#16817](#), amitkuma)
- msg: Initializing uninitialized members MOSDALive ([pr#16816](#), amitkuma)
- msg: make listen backlog an option, increase from 128 to 512 ([issue#20330](#), [pr#15743](#), Haomai Wang)
- msg: messages: coverity fixes ([pr#13473](#), Kefu Chai)
- msg: msg/async: avoid atomic variable overhead ([pr#12809](#), Wei Jin)
- msg: msg/async: cleanup code ([pr#13304](#), Jianpeng Ma)
- msg: msg/async: cleanups ([pr#12832](#), Wei Jin)
- msg: msg/async: fix file description leak in NetHandler ([pr#13271](#), liuchang0812)

- msg: msg/async: increase worker reference with local listen table enabled backend ([issue#20390](#), [pr#15897](#), Haomai Wang)
- msg: msg/async: Lower down the AsyncMessenger's standby warning from debug ([pr#15242](#), Pan Liu)
- msg: msg/AsyncMessenger: remove unused method ([pr#10125](#), Michal Jarzabek)
- msg: msg/async/net\_handler: errno should be stored before calling next function ([pr#14985](#), Zhou Zhengping)
- msg: msg/async/rdma: check if exp verbs avail ([pr#13391](#), Oren Duer, Adir Lev)
- msg: msg/async/rdma: check if fin message completed ([pr#15624](#), Alexander Mikheev, Adir Lev)
- msg: msg/async/rdma: Data path fixes ([pr#15903](#), Adir lev)
- msg: msg/async/rdma: Debug prints for ibv\* ([pr#14249](#), Amir Vadai)
- msg: msg/async/rdma: Device::last\_poll\_dev must be positive ([pr#14250](#), Amir Vadai)
- msg: msg/async/rdma: Fix broken compilation ([pr#13603](#), Sarit Zubakov)
- msg: msg/async/rdma: Fix memory leak of OSD ([pr#13101](#), Sarit Zubakov)
- msg: msg/async/rdma: fix outstanding queuepair when destruct RDMAStack ([pr#13905](#), Haomai Wang)
- msg: msg/async/rdma: fix RoCE v2 deafult value ([pr#12648](#), Adir Lev, Oren Duer)
- msg: msg/async/rdma: Fix small memory leaks detected by valgrind ([pr#14288](#), Amir Vadai)
- msg: msg/async/rdma: handle buffers after close msg ([pr#15749](#), DanielBar-On, Alexander Mikheev, Adir Lev)
- msg: msg/async/rdma: move active\_queue\_pairs perf counter dec to polling ([pr#13716](#), DanielBar-On)
- msg: msg/async/rdma: Print error only on ENOMEM ([pr#13538](#), Sarit Zubakov)
- msg: msg/async/rdma: RDMA-CM, Pass specific ConnMgr info in constructor ([pr#14409](#), Amir Vadai)
- msg: msg/async/rdma: register buffer as continuous ([pr#15967](#), Adir Lev)
- msg: msg/async/rdma: remove assert from ibv\_dealloc\_pd in ProtectionDomain ([pr#15832](#), DanielBar-On)
- msg: msg/async/rdma: update destructor message ([pr#13539](#), Sarit Zubakov)

- msg: msg/async/rdma: zero wqe inline ([pr#13392](#), Adir Lev)
- msg: msg/async: remove false alert “assert” ([pr#15288](#), Haomai Wang)
- msg: msg/async: remove useless close function ([pr#13286](#), liuchang0812)
- msg: msg/async: rm nonused thread variable in posixworker ([pr#12777](#), Wei Jin)
- msg: msg/async: use auto iterator having more simple code and good performance ([pr#16524](#), dingdangzhang)
- msg: msg/Messenger.cc: add std::move ([pr#9760](#), Michal Jarzabek)
- msg: msg/MOSDOpReply: fix missing trace decode ([pr#15999](#), Yan Jun)
- msg: msg/RDMA: Fix broken compilation due to new argument in net.connect() ([pr#13096](#), Amir Vadai)
- msg: msg/simple: Remove dead code in pipe.cc ([issue#12684](#), [pr#12601](#), Rishabh Kumar)
- msg: msg/simple: use my addr when setting sock priority ([issue#19801](#), [pr#14878](#), Kefu Chai)
- msg: no need to pass supported features to Messenger::Policy ctor ([pr#13785](#), Sage Weil)
- msg: QueueStrategy::wait() joins all threads ([issue#20534](#), [pr#16194](#), Casey Bodley)
- msg: Remove unused variable perf\_counter in RDMAStack ([pr#16783](#), amitkuma)
- msg: Revert the change from assert(0)-> ceph\_abort() where is not applicable ([pr#12930](#), Dave Chen)
- msg: src/msg/async/AsyncConnect.cc: Use of sizeof() on a Pointer Type ([pr#14773](#), Svyatoslav)
- msg: src/msg/async: Update fix broken compilation for Posix ([pr#14336](#), Sarit Zubakov)
- msg: src/msg/simple/Pipe.cc: Fix the inclusion of ‘}’ ([pr#14843](#), Willem Jan Withagen)
- os/bluestore: print leaked extents to debug output ([pr#17303](#), Sage Weil)
- osd: add asock command to dump the scrub queue ([issue#17861](#), [pr#12728](#), liuchang0812)
- osd: add default\_device\_class to metadata ([pr#16634](#), Neha Ojha)
- osd: add dump filter for tracked ops ([pr#16561](#), Yan Jun)

- osd: add “heap \*” admin command ([issue#15475](#), [pr#13073](#), Jesse Williamson)
- osd: adding PerfCounters for backoff throttle ([pr#13017](#), Chuanhong Wang)
- osd: add is\_split check before \_start\_split ([pr#13307](#), song baisen)
- osd: add override in headers files ([pr#13962](#), liuchang0812)
- osd: add override in osd subsystem ([issue#18922](#), [pr#13439](#), liuchang0812)
- osd: Add recovery sleep configuration option for HDDs and SSDs ([pr#16328](#), Neha Ojha)
- osd: add snap trim reservation and re-implement osd\_snap\_trim\_sleep ([pr#13594](#), Samuel Just)
- osd: adjust osd\_min\_pg\_log\_entries ([issue#21026](#), [pr#17202](#), J. Eric Ivancich)
- osd: allow client throttler to be adjusted on-fly, without restart ([issue#18791](#), [pr#13213](#), Piotr Dałek)
- osd: bail from \_committed\_osd\_maps inside osd\_lock ([issue#20273](#), [pr#15710](#), Sage Weil)
- osd: Calculate degraded and misplaced more accurately ([issue#18619](#), [pr#13031](#), David Zafman)
- osd: change a few messages at level 0 and 1; change default level to 1/5 ([pr#13407](#), Sage Weil)
- osd: Check for and automatically repair object info soid during scrub ([issue#20471](#), [pr#16052](#), David Zafman)
- osd: check fsid is normal before osd mkfs ([pr#13898](#), song baisen)
- osd: check queue\_transaction return value ([pr#15873](#), zhanglei)
- osd: Check snapset for validity when selecting authoritative shard ([issue#20186](#), [issue#18409](#), [pr#15559](#), David Zafman)
- osd: Check whether journal is rotational or not ([pr#16614](#), Neha Ojha)
- osd: clarify REQUIRE\_LUMINOUS error message ([pr#13363](#), Josh Durgin)
- osd: clean nonused work queue ([pr#14990](#), Wei Jin)
- osd: Cleanup-Updated OSDMap.cc with C++11 style range-for loops ([pr#14381](#), Jos Collin)
- osd: cleanup: use string & to avoid unnecessary copy ([pr#12336](#), Yunchuan Wen)
- osd: clear\_queued\_recovery() in on\_shutdown() ([issue#20432](#), [pr#16093](#), Kefu Chai)

- osd: cmpext operator should ignore -ENOENT on read ([pr#16622](#), Jason Dillaman)
- osd: combine conditional statements ([pr#16391](#), Yan Jun)
- osd: combine unstable stats with info.stats when publish stats to osd ([pr#14060](#), Mingxin Liu)
- osd: compact osd feature ([issue#19592](#), [pr#16045](#), liuchang0812)
- osd: condition object\_info\_t encoding on required (not up) features ([issue#18644](#), [pr#13114](#), Sage Weil)
- osd: constify OpRequest::get\_req(); fix a few cases of operator<< vs mutated message races ([pr#13545](#), Sage Weil)
- osd: correct comment of perfcounter cached\_crc in code ([pr#13256](#), lvshuhua)
- osd: correct epoch setting of osd boot msg ([pr#12623](#), Mingxin Liu)
- osd: correct the func name in execute\_ctx() log messages ([pr#13582](#), Gu Zhongyan)
- osd: Corrupt objects stop snaptrim and mark pg snaptrim\_error ([issue#13837](#), [pr#15635](#), David Zafman)
- osd: debug con in ms\_handle\_connect ([pr#13540](#), Sage Weil)
- osd: do not send ENXIO on misdirected op by default ([issue#18751](#), [pr#13206](#), Sage Weil)
- osd: do not send pg\_created unless luminous ([issue#20785](#), [pr#16677](#), Kefu Chai)
- osd: do not try to boot until we've seen the first osdmap ([pr#15732](#), Sage Weil)
- osd: do not try to set device class before luminous ([issue#20850](#), [pr#16706](#), Josh Durgin)
- osd: don't leak pgrefs or reservations in SnapTrimmer ([issue#19931](#), [pr#15214](#), Greg Farnum)
- osd: don't share osdmap with objecter when preboot ([issue#15025](#), [pr#13946](#), Mingxin Liu)
- osd: don't use ORDERSNAP for flush; always request/send ondisk ack ([issue#18961](#), [pr#13570](#), Samuel Just)
- osd: drop support for listing objects at a given snap ([pr#13398](#), Sage Weil)
- osd: dump the field name of object watchers and cleanups ([pr#15946](#), Yan Jun)
- osd: EC read handling: don't grab an objectstore error to use as the read error ([pr#16663](#), David Zafman)

- osd: eliminate snapdir objects and move clone snaps vector into SnapSet ([pr#13610](#), Sage Weil)
- osd: Execute crush\_location\_hook when configured in ceph.conf ([pr#15951](#), Wido den Hollander)
- osd: \_exit() instead of exit() for failure injection ([issue#18372](#), [pr#12726](#), Sage Weil)
- osd: extend OMAP\_GETKEYS and GETVALS to include a 'more' output field ([pr#12950](#), Sage Weil)
- osd: fall back to failsafe threshold if osdmap doesn't set [near]full ([pr#14004](#), Sage Weil)
- osd: faster dispatch ([pr#13343](#), Sage Weil)
- osd: fix a couple bugs with persisting the missing set when it contains deletes ([issue#20704](#), [pr#16459](#), Josh Durgin)
- osd: fix argument-dependent lookup of swap() ([pr#15124](#), Casey Bodley)
- osd: fix a signed/unsigned warning in PG ([pr#13922](#), Greg Farnum)
- osd: fix comments about pg refs and lock ([pr#14279](#), tang.jin)
- osd: fix coverity warning for uninitialized members ([pr#12724](#), Li Wang)
- osd: fix func name in log produced by handle\_pg\_peering\_evt() ([pr#13801](#), xie xingguo)
- osd: fix occasional MOSDMap leak ([issue#18293](#), [pr#14558](#), Sage Weil)
- osd: fix OpRequest and tracked op dump information ([pr#16504](#), Yan Jun)
- osd: fix past\_intervals base case by adding epoch\_pool\_created to pg\_history\_t ([issue#19877](#), [pr#14989](#), Sage Weil)
- osd: fix pg ref leaks when osd shutdown ([issue#20684](#), [pr#16408](#), Yang Honggang)
- osd: fix some osd beacon bugs ([pr#14274](#), Sage Weil)
- osd: fix stat sum update of recovery pushing ([pr#13328](#), Zhiqiang Wang)
- osd: fix the setting of soid in sub\_op\_push ([pr#13353](#), Zhiqiang Wang)
- osd: fix typo in comment ([pr#13061](#), Gu Zhongyan)
- osd: Fix useless MAX(0, unsigned) to prevent out of wack misplaced ([issue#18718](#), [pr#13164](#), David Zafman)
- osd: have clients resend ops on pg split ([pr#13235](#), Sage Weil)

- osd: hdd vs ssd defaults for osd op thread pool ([pr#15422](#), Sage Weil)
- osd: heartbeat with packets large enough to require working jumbo frames ([issue#20087](#), [pr#15535](#), Greg Farnum)
- osd: Implement asynchronous recovery sleep ([pr#15212](#), Neha Ojha)
- osd: Implement asynchronous scrub sleep ([issue#19497](#), [pr#14886](#), Brad Hubbard)
- osd: Implement peering state timing ([pr#14627](#), Brad Hubbard)
- osd: improve error message when FileStore op fails due to EPERM ([issue#18037](#), [pr#12181](#), Nathan Cutler)
- osd: initialize waiting\_for\_pg\_osdmap on startup ([issue#20748](#), [pr#16535](#), Sage Weil)
- osd: kill all remaining MOSDSubOp users ([pr#13401](#), Sage Weil)
- osd: kill sortbitwise ([pr#13321](#), Sage Weil)
- osd: Log audit ([pr#16281](#), Brad Hubbard)
- osd: make ec overwrites ready to use ([pr#14496](#), Josh Durgin)
- osd: moved OpFinisher logic from OSDOp to OpContext ([issue#20783](#), [pr#16617](#), Jason Dillaman)
- osd: Move scrub sleep timer to osdservice ([issue#19986](#), [pr#15217](#), Brad Hubbard)
- osd: never send rados ack (only commit) ([pr#12451](#), Sage Weil)
- osd: new op for calculating an extent checksum ([pr#14256](#), Jason Dillaman)
- osd: objclass sdk ([pr#14723](#), Neha Ojha)
- osd: Object level shard errors are tracked and used if no auth available ([issue#20089](#), [pr#15397](#), David Zafman)
- osd: On EIO from read recover the primary replica from another copy ([issue#18165](#), [pr#14760](#), David Zafman)
- osd: osdc/ObjectCacher: use state instead of get\_state() ([pr#12544](#), huangjun)
- osd: osdc/Objecter: more constness ([pr#14819](#), Kefu Chai)
- osd: osdc: silence warning from -Wsign-compare ([pr#14729](#), Jos Collin)
- osd: osd does not using MPing Messages, do not include unused include ([pr#15833](#), linbing)
- osd: osd/OSDMap.cc: check if osd is out in subtree\_type\_is\_down ([issue#19989](#), [pr#15250](#), Neha Ojha)

- osd: osd/OSDMap: require OSD features only of OSDs ([issue#18831](#), [pr#13275](#), Ilya Dryomov)
- osd: osd/PrimaryLogPG: nullptr not NULL ([pr#13973](#), Shinobu Kinjo)
- osd: 'osd tree in|out|up|down' to filter tree results ([pr#15294](#), Sage Weil)
- osd: os/kstore: some error handling ([pr#13960](#), wangzhengyong)
- osd/PGBackend: delete reply if fails to complete delete request ([issue#20913](#), [pr#17233](#), Kefu Chai)
- osd: pg: be more careful with locking around forced pg recovery ([issue#20808](#), [pr#16712](#), Greg Farnum)
- osd: pglog trimming fixes ([pr#12882](#), Zhiqiang Wang)
- osd: pglog: with config, don't assert in the presence of stale diverg... ([issue#17916](#), [pr#14648](#), Greg Farnum)
- osd: pg-remap -> pg-upmap ([pr#14556](#), Sage Weil)
- osd: populate last\_epoch\_split during build\_initial\_pg\_history ([issue#20754](#), [pr#16519](#), Sage Weil)
- osd: Preserve OSDOp information for historic ops ([pr#15265](#), Guo-Fu Tseng)
- osd: PrimaryLogPG, PGBackend: complete callback even if interval changes ([issue#20747](#), [pr#16536](#), Josh Durgin)
- osd: print pg\_info\_t::purged\_snaps as array, not string ([issue#18584](#), [pr#14217](#), liuchang0812)
- osd: process deletes during recovery instead of peering ([issue#19971](#), [pr#15952](#), Josh Durgin)
- osd: put osdmap in mempool ([pr#14780](#), Sage Weil)
- osd: reduce buffer pinning from EC entries ([pr#15120](#), Sage Weil)
- osd: reduce map cache size ([pr#15292](#), Sage Weil)
- osd: reduce rados\_max\_object\_size from 100 GB -> 128 MB ([pr#15520](#), Sage Weil)
- osd: remove copy-get-classic ([pr#13547](#), Sage Weil)
- osd: remove sortbitwise thrashing ([pr#13296](#), Sage Weil)
- osd: renamed the new vector name in OSDMap::build\_simple\_crush\_map\_from\_conf ([pr#14583](#), Jos Collin)
- osd: rename osd -> osd\_pglog; include pglog-related bufferlists ([pr#15531](#), Sage

Weil)

- osd: rephrase “wrongly marked me down” clog message ([pr#16365](#), John Spray)
- osd: replace object\_info\_t::operator=() with decode() ([pr#13938](#), tang.jin)
- osd: ReplicatedBackend::prep\_push() remove redundant variable assignments ([pr#14817](#), Jin Cai)
- osd: restart boot process if waiting for luminous mons ([issue#20631](#), [pr#16341](#), Sage Weil)
- osd: Return correct osd\_objectstore in OSD metadata ([issue#18638](#), [pr#13072](#), Wido den Hollander)
- osd: Return early on shutdown ([issue#19900](#), [pr#15345](#), Brad Hubbard)
- osd: Reverse order of op\_has\_sufficient\_caps and do\_pg\_op ([issue#19790](#), [pr#15354](#), Brad Hubbard)
- osd: sched\_scrub() lock pg only if all scrubbing conditions are fulfilled ([pr#14968](#), Jin Cai)
- osd: scrub\_to specifies clone ver, but transaction include head write... ([issue#20041](#), [pr#16404](#), David Zafman)
- osd: silence warning from -Wint-in-bool-context ([pr#16744](#), Jos Collin)
- osd: simplify past\_intervals representation ([pr#14444](#), Samuel Just, Sage Weil)
- osd: small clear up and optimize on \_recover\_now and should\_share\_map function ([pr#13476](#), song baisen)
- osd: stop mgrc earlier in shutdown() ([issue#19638](#), [pr#14904](#), Kefu Chai)
- osd: stop MgrClient callbacks on shutdown ([issue#19638](#), [pr#14896](#), Sage Weil)
- osd: strip pglog op name ([pr#14764](#), liuchang0812)
- osd: support cmpext operation on EC-backed pools ([pr#15693](#), Zhengyong Wang, Jason Dillaman)
- osd: support dumping long ops ([pr#13019](#), Zhiqiang Wang)
- osd: switch filestore to default to rocksdb ([pr#14814](#), Neha Ojha)
- osd: tag fast dispatch messages with min\_epoch ([pr#13681](#), Sage Weil)
- osd: take PGRef for recovery sleep wakeup event ([issue#20226](#), [pr#15582](#), Sage Weil)
- osd: the condition of last epoch <= superblock.newest\_map epoch has been check

- twice ([pr#15590](#), linbing)
- osd: the osd should not share map with others when it is in stopping state ([pr#13668](#), song baisen)
- osd: unlock sdata\_op\_ordering\_lock with sdata\_lock hold to avoid miss... ([pr#15891](#), Ming Lin)
- osd: use append(bufferlist &) to avoid unnecessary copy ([pr#12272](#), Yunchuan Wen)
- osd: use separate waitlist for scrub ([pr#13136](#), Sage Weil)
- osd: various changes for preventing internal ENOSPC condition ([issue#16878](#), [pr#13425](#), David Zafman)
- osd: we know the definite epoch of marking down ([pr#13121](#), Mingxin Liu)
- osd: when osd is not in failure\_pending, we don't need to get osd inst from osdmap ([pr#15558](#), linbing)
- osd: When scrub finds an attr error mark shard inconsistent ([issue#20089](#), [pr#15368](#), David Zafman)
- osd: zipkin tracing ([pr#14305](#), Sage Weil, Marios-Evaggelos Kogias, Victor Araujo, Casey Bodley, Andrew Shewmaker, Chendi.Xue)
- performance: buffer, osd: add missing crc cache miss perf counter ([pr#14957](#), Piotr Dałek)
- performance: common/config\_opts.h: Lower HDD throttle cost ([pr#15485](#), Mark Nelson)
- performance: crc32c: optimize aarch64 crc32c implementation ([pr#12977](#), wei xiao)
- performance: denc: add need\_contiguous to denc\_traits ([pr#15224](#), Kefu Chai)
- performance: osd, messenger, librados: lttng oid tracing ([pr#12492](#), Anjaneya Chagam)
- performance: osd/PG.cc: loop invariant code motion ([pr#12720](#), Li Wang)
- performance: osd/ReplicatedBackend: do not set omap header if it is empty ([pr#12612](#), fang yuxiang)
- performance,rgw: rgw\_file: permit dirent offset computation ([pr#16275](#), Matt Benjamin)
- pybind: better error msg ([pr#14497](#), Kefu Chai)
- pybind: cephfs should be built without librados / python-rados ([pr#13431](#), Kefu Chai)

- pybind: ceph.in: Check return value when connecting ([pr#16130](#), Douglas Fuller)
- pybind: ceph-rest-api: Various REST API fixes ([pr#15910](#), Wido den Hollander)
- pybind: conditional compile the linux specific constant ([pr#12198](#), Kefu Chai)
- pybind: fix docstring for librbd Python binding ([pr#13977](#), runsisi)
- pybind: fix open flags calculation ([issue#19890](#), [pr#15018](#), "Yan, Zheng")
- pybind: pybind/ceph\_argparse: fix empty string check ([issue#20135](#), [pr#15500](#), Sage Weil)
- pybind: pybind/ceph\_daemon.py: fix Termsize.update ([pr#15253](#), Kefu Chai)
- pybind: pybind/ceph\_daemon: use small chunk for recv ([pr#13804](#), Xiaoxi Chen)
- pybind: pybind/mgr/dashboard: fix get kernel\_version error ([pr#16094](#), Peng Zhang)
- pybind: pybind/mgr/restful: fix typo ([pr#16560](#), Nick Erdmann)
- pybind: pybind/mgr/restful: use list to pass hooks to create a Pecan instance ([issue#20258](#), [pr#15646](#), Kefu Chai)
- pybind: pybind/rados: avoid call free() on invalid pointer ([pr#15159](#), Mingxin Liu)
- pybind: pybind/rados: use new APIs instead of deprecated ones ([pr#16684](#), Kefu Chai)
- pybind, rbd: pybind/rbd: OSError should be picklable ([issue#20223](#), [pr#15574](#), Jason Dillaman)
- pybind: restore original API for backwards compatibility ([issue#20421](#), [pr#15932](#), Jason Dillaman)
- pybind: support mon target in pybind ([pr#15409](#), liuchang0812)
- qa: fix POOL\_APP\_NOT\_ENABLED warning in krbd:unmap suite ([pr#17192](#), Ilya Dryomov)
- rbd: add default note info to size (create and resize) ([pr#15561](#), Zheng Yin)
- rbd: add error prompt when input command 'snap set limit' is incomplete ([pr#12945](#), Tang Jin)
- rbd: additional validation for 'bench' optional parameters ([pr#12697](#), Yunchuan Wen)
- rbd: bench-write should return error if io-size >= 4G ([issue#18422](#), [pr#12864](#), Gaurav Kumar Garg)
- rbd: cleanup: fix the typo in namespace comment ([pr#12858](#), Dongsheng Yang)

- rbd: cleanup: rbd: fix a typo in comment ([pr#14049](#), Dongsheng Yang)
- rbd: cls\_rbd: default initialize snapshot namespace for legacy clients ([issue#19413](#), [pr#14903](#), Jason Dillaman)
- rbd: cls/rbd: silence warning from -Wunused-variable ([pr#16670](#), Yan Jun)
- rbd: cls/rbd: trash\_list should be iterable ([issue#20643](#), [pr#16372](#), Jason Dillaman)
- rbd: common/bit\_vector: utilize deep-copy during data decode ([issue#19863](#), [pr#15017](#), Jason Dillaman)
- rbd: correct coverity warnings ([pr#12954](#), Jason Dillaman)
- rbd: correct issues with image importing ([pr#14401](#), Jason Dillaman)
- rbd: demote/promote all mirrored images in a pool ([issue#18748](#), [pr#13758](#), Jason Dillaman)
- rbd: destination pool should be source pool if it is not specified ([issue#18326](#), [pr#13189](#), Gaurav Kumar Garg)
- rbd: do not attempt to load key if auth is disabled ([issue#19035](#), [pr#16024](#), Jason Dillaman)
- rbd: Drop unused member variable reopen in C\_OpenComplete ([pr#16729](#), amitkuma)
- rbd: enable rbd on FreeBSD (without KRBD) ([pr#12798](#), Willem Jan Withagen)
- rbd: error out if import image format failed ([pr#13957](#), wangzhengyong)
- rbd: fixed coverity ‘Argument cannot be negative’ warning ([pr#16686](#), amitkuma)
- rbd: fix typo in Kernel.cc ([issue#19273](#), [pr#13983](#), Gaurav Kumar Garg)
- rbd: ‘image-meta remove’ for missing key does not return error ([issue#16990](#), [pr#16393](#), PCzhangPC)
- rbd: import-diff should discard any zeroed extents ([pr#14445](#), Jason Dillaman)
- rbd: import needs to sanity check auto-generated image name ([issue#19128](#), [pr#14754](#), Mykola Golub)
- rbd: import real thin-provision image ([issue#15648](#), [pr#12883](#), yaoning, Ning Yao)
- rbd: info command should indicate if parent is in trash ([pr#14875](#), Jason Dillaman)
- rbd: introduce v2 format for rbd export/import ([issue#13186](#), [pr#10487](#), Dongsheng Yang)

- rbd: journal: don't hold future lock during assignment ([issue#18618](#), [pr#13033](#), Jason Dillaman)
- rbd: journal: stop processing removal after error ([issue#18738](#), [pr#13193](#), Jason Dillaman)
- rbd: luminous: librbd: default localize parent reads to false ([issue#20941](#), [pr#16899](#), Jason Dillaman)
- rbd: luminous: librbd: remove consistency group rbd cli and API support ([pr#16875](#), Jason Dillaman)
- rbd: luminous: rbd-ggate: tool to map images on FreeBSD via GEOM Gate ([pr#16895](#), Mykola Golub)
- rbd: luminous: rbd-mirror: align use of uint64\_t in service\_daemon::AttributeType ([pr#16948](#), James Page)
- rbd: luminous: rbd-mirror: simplify notifications for image assignment ([issue#15764](#), [pr#16878](#), Jason Dillaman)
- rbd: luminous: rbd: parallelize rbd ls -l ([pr#16921](#), Piotr Dałek)
- rbd: make it more understandable when adding peer returns error ([pr#16313](#), songweibin)
- rbd-mirror: add support for active/pассиве daemon instances ([issue#17018](#), [issue#17019](#), [issue#17020](#), [pr#12948](#), Mykola Golub)
- rbd-mirror: assertion failure during pool replayer shut down ([issue#20644](#), [pr#16704](#), Jason Dillaman)
- rbd-mirror: avoid processing new events after stop requested ([issue#18441](#), [pr#12837](#), Jason Dillaman)
- rbd-mirror: check remote image mirroring state when bootstrapping ([issue#18447](#), [pr#12820](#), Mykola Golub)
- rbd-mirror: coordinate image syncs with leader ([issue#18789](#), [pr#14745](#), Mykola Golub)
- rbd-mirror: delayed replication support ([issue#15371](#), [pr#11879](#), Mykola Golub)
- rbd-mirror: deleting a snapshot during sync can result in read errors ([issue#18990](#), [pr#13568](#), Jason Dillaman)
- rbd-mirror: ensure missing images are re-synced when detected ([issue#19811](#), [pr#14945](#), Jason Dillaman)
- rbd-mirror: failover and fallback of unmodified image results in split-brain ([issue#19858](#), [pr#14963](#), Jason Dillaman)

- rbd-mirror: guard the deletion of non-primary images ([pr#16398](#), Jason Dillaman)
- rbd-mirror: ignore permission errors on rbd\_mirroring object ([issue#20571](#), [pr#16264](#), Jason Dillaman)
- rbd-mirror: image deletions should be handled by assigned instance ([pr#14832](#), Jason Dillaman)
- rbd-mirror: initialize timer context pointer to null ([pr#16603](#), Jason Dillaman)
- rbd-mirror: InstanceWatcher watch/notify stub for leader/follower RPC ([issue#18783](#), [pr#13312](#), Mykola Golub)
- rbd-mirror: lock loss during sync should wait for in-flight copies ([pr#15532](#), Jason Dillaman)
- rbd-mirror: permit release of local image exclusive lock after force promotion ([issue#18963](#), [pr#15140](#), Jason Dillaman)
- rbd-mirror: pool watcher should track mirror uuid ([pr#14240](#), Jason Dillaman)
- rbd-mirror: remove tracking of image names from pool watcher ([pr#14712](#), Jason Dillaman)
- rbd-mirror: replace remote pool polling with add/remove notifications ([issue#15029](#), [pr#12364](#), Jason Dillaman)
- rbd-mirror: resolve admin socket path names collision ([issue#19907](#), [pr#15048](#), Mykola Golub)
- rbd-mirror: separate ImageReplayer handling from Replayer ([issue#18785](#), [pr#13803](#), Mykola Golub)
- rbd-mirror: Set the data pool correctly when creating images ([issue#20567](#), [pr#17023](#), Adam Wolfe Gordon)
- rbd-mirror: track images via global image id ([pr#13416](#), Jason Dillaman)
- rbd: modified some commands' description into imperative sentence ([pr#16694](#), songweibin)
- rbd-nbd: check /sys/block/nbdX/size to ensure kernel mapped correctly ([issue#18335](#), [pr#13229](#), Mykola Golub)
- rbd-nbd: clean up the doc and help information ([pr#14146](#), Pan Liu)
- rbd-nbd: create admin socket only for map command ([issue#17951](#), [pr#12433](#), Pan Liu)
- rbd-nbd: display pool/image/snap information in list output ([pr#15317](#), Pan Liu)
- rbd-nbd: don't ignore -read-only option in BLKROSET ioctl ([pr#13944](#), Pan Liu)

- rbd-nbd: ensure unmap returns error code ([pr#15593](#), guojiannan, chenfangxian)
- rbd-nbd: fix a typo “moudle” ([pr#13652](#), Pan Liu)
- rbd-nbd: fix typo in comment ([pr#14034](#), Pan Liu)
- rbd-nbd: no need to check image format any more ([pr#13389](#), Mykola Golub)
- rbd-nbd: relax size check for newer kernel versions ([issue#19871](#), [pr#14976](#), Mykola Golub)
- rbd-nbd: remove debug messages from do\_unmap ([pr#14253](#), Pan Liu)
- rbd-nbd: s/cpp\_error/cpp\_strerror/ to fix FTBFS ([pr#14223](#), Kefu Chai)
- rbd-nbd: support signal handle for SIGHUP, SIGINT and SIGTERM ([issue#19349](#), [pr#14079](#), Pan Liu)
- rbd-nbd: update size only when NBD\_SET\_SIZE successful ([pr#14005](#), Pan Liu)
- rbd-nbd: warn when kernel parameters are ignored ([issue#19108](#), [pr#13694](#), Pan Liu)
- rbd: prevent adding multiple mirror peers to a single pool ([issue#19256](#), [pr#13919](#), Jason Dillaman)
- rbd: properly decode features when using image name optional ([issue#20185](#), [pr#15492](#), Jason Dillaman)
- rbd: pybind/rbd: add image metadata methods ([issue#19451](#), [pr#14463](#), Mykola Golub)
- rbd: pybind/rbd: fix crash if more than 1024 images in trash bin ([pr#15134](#), runnsisi)
- rbd: rbd/bench: add notes of default values, it's easy to use ([pr#14762](#), Zheng Yin)
- rbd: rbd/bench: fix write gaps when doing sequential writes with io-threads > 1 ([pr#15206](#), Igor Fedotov)
- rbd: rbd, librbd: migrate atomic\_t to std::atomic ([pr#14656](#), Jesse Williamson)
- rbd: rbd-mirror A/A: leader should track up/down rbd-mirror instances ([issue#18784](#), [pr#13571](#), Mykola Golub)
- rbd: rbd-mirror A/A: proxy InstanceReplayer APIs via InstanceWatcher RPC ([issue#18787](#), [pr#13978](#), Mykola Golub)
- rbd: recognize exclusive option ([pr#14785](#), Ilya Dryomov)
- rbd: removed hardcoded default pool ([pr#15518](#), Jason Dillaman)
- rbd: remove direct linking to static boost libraries ([pr#12962](#), Jason Dillaman)

- rbd: removed spurious error message from mirror pool commands ([pr#14935](#), Jason Dillaman)
- rbd: remove unused condition within group action handler ([pr#12723](#), Gaurav Kumar Garg)
- rbd,rgw,tools: tools/rbd, rgw: Removed unreachable returns ([pr#16308](#), Jos Collin)
- rbd: spell out image features unsupported by the kernel ([issue#19095](#), [pr#13812](#), Ilya Dryomov)
- rbd: stop indefinite thread waiting in krbd udev handling ([issue#17195](#), [pr#14051](#), Spandan Kumar Sahu)
- rbd: test: fix rbd unit test cases w/ striping feature ([issue#18888](#), [pr#13196](#), Venky Shankar)
- rbd,tests: luminous: qa/workunits/rbd: use command line option to specify watcher asok ([issue#20954](#), [pr#16946](#), Mykola Golub)
- rbd,tests: luminous: test/librbd: fix race condition with OSD map refresh ([issue#20918](#), [pr#16903](#), Jason Dillaman)
- rbd,tests: qa: add workunit to test krbd data-pool support ([pr#13482](#), Ilya Dryomov)
- rbd,tests: qa: integrate OpenStack ‘gate-tempest-dsvm-full-devstack-plugin-ceph’ ([issue#18594](#), [pr#13158](#), Jason Dillaman)
- rbd,tests: qa: krbd\_data\_pool.sh: account for rbd\_info metadata object ([pr#14631](#), Ilya Dryomov)
- rbd,tests: qa: krbd discard/zeroout tests ([pr#15388](#), Ilya Dryomov)
- rbd,tests: qa: krbd write-after-checksum tests ([pr#14836](#), Ilya Dryomov)
- rbd,tests: qa/suites/krbd: unmap subsuite needs straw buckets ([pr#15290](#), Ilya Dryomov)
- rbd,tests: qa/suites/rbd: restrict python memcheck validation to CentOS ([pr#15923](#), Jason Dillaman)
- rbd,tests: qa/tasks/qemu: update default image url after ceph.com redesign ([issue#18542](#), [pr#12953](#), Jason Dillaman)
- rbd,tests: qa/tasks/rbd\_fio: bump default fio version to 2.21 ([pr#16656](#), Ilya Dryomov)
- rbd,tests: qa/tasks: rbd-mirror daemon not properly run in foreground mode ([issue#20630](#), [pr#16340](#), Jason Dillaman)
- rbd,tests: qa: thrash tests for backoff and upmap ([pr#16428](#), Ilya Dryomov)

- rbd,tests: qa: update krbd\_data\_pool.sh to match the new rados ls behavior ([pr#15594](#), Ilya Dryomov)
- rbd,tests: qa/workunits: adjust path to ceph-helpers.sh ([pr#16599](#), Sage Weil)
- rbd,tests: qa/workunits: corrected issues with RBD cli test ([pr#14460](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: diff.sh failed removing nonexistent file ([pr#14482](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: increased trash deferment period ([pr#14846](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: resolve potential rbd-mirror race conditions ([issue#18935](#), [pr#13421](#), Jason Dillaman)
- rbd,tests: qa/workunits/rbd: test data pool is mirrored correctly ([pr#17077](#), Mykola Golub)
- rbd,tests: qa/workunits/rbd: tweak rbd-mirror config to speed up testing ([pr#13228](#), Mykola Golub)
- rbd,tests: qa/workunits: switch to OpenStack Ocata release for RBD testing ([pr#14465](#), Jason Dillaman)
- rbd,tests: test: correct language mode in file headers ([pr#12924](#), Jason Dillaman)
- rbd,tests: test: fix compile warning in ceph\_test\_cls\_rbd ([pr#15919](#), Jason Dillaman)
- rbd,tests: test: fix failing rbd devstack teuthology test ([pr#15956](#), Jason Dillaman)
- rbd,tests: test/librados\_test\_stub: fixed cls\_cxx\_map\_get\_keys/vals return value ([issue#19597](#), [pr#14484](#), Jason Dillaman)
- rbd,tests: test/librbd: add break\_lock test ([pr#12842](#), Mykola Golub)
- rbd,tests: test/librbd/CMakeLists.txt: ceph\_test\_librbd\_fsx requires linux includes/libs ([pr#13630](#), Willem Jan Withagen)
- rbd,tests: test/librbd/fsx: Add break in case OP\_WRITESAME and OP\_COMPARE\_AND\_WRITE ([pr#16742](#), Luo Kexue)
- rbd,tests: test/librbd: move tests using non-public api to internal ([pr#13806](#), Venky Shankar)
- rbd,tests: test/librbd/test\_librbd.cc: set \*features even if RBD\_FEATURES is unset ([issue#19865](#), [pr#14965](#), Dan Mick)
- rbd,tests: test/librbd/test\_notify.py: don't disable feature in slave

([issue#19716](#), [pr#14751](#), Mykola Golub)

- rbd,tests: test/librbd: unit tests cleanup ([pr#15113](#), Mykola Golub)

- rbd,tests: test: rbd master/slave notify test should test active features ([issue#19692](#), [pr#14638](#), Jason Dillaman)
- rbd,tests: test/rbd\_mirror: race in TestMockInstanceWatcher on destroy ([pr#14453](#), Mykola Golub)
- rbd,tests: test/rbd\_mirror: race in TestMockLeaderWatcher.AcquireError ([issue#19405](#), [pr#14741](#), Mykola Golub)
- rbd,tests: test: remove hard-coded image name from RBD metadata test ([issue#19798](#), [pr#14848](#), Jason Dillaman)
- rbd,tests: test: support blacklisting within librados\_test\_stub ([pr#13737](#), Jason Dillaman)
- rbd,tests: test/unittest\_librbd: remove unused variables ([pr#15720](#), shiqi)
- rbd,tests: test: use librados API to retrieve config params ([issue#18617](#), [pr#13076](#), Jason Dillaman)
- rbd,tools: rbdmap: consider /etc/ceph/rbdmap when unmapping images ([issue#18884](#), [pr#13361](#), David Disseldorp)
- rbd,tools: tools/rbd\_mirror: initialize non-static class member m\_do\_resync in ImageReplayer ([pr#15889](#), Jos Collin)
- rbd,tools: tools/rbd\_nbd: add --version show support ([pr#16254](#), Jin Cai)
- rbd: use concurrent writes for imports ([issue#19034](#), [pr#13782](#), Venky Shankar)
- rbd: use min<uint64\_t>() explicitly ([issue#18938](#), [pr#14202](#), Kefu Chai)
- rbd: validate pool and snap name optionals ([issue#14535](#), [pr#13836](#), Gaurav Kumar Garg)
- rbd: warning, 'devno' may be used uninitialized in this function ([pr#14271](#), Jos Collin)
- rbd: When Ceph cluster becomes full, should allow user to remove rbd ... ([pr#12627](#), Pan Liu)
- rdma: msg/async: Postpone bind if network stack is not ready ([pr#14414](#), Amir Vadai, Haomai Wang)
- rdma: msg/async/rdma: Add DSCP support ([pr#15484](#), Sarit Zubakov)
- rdma: msg/async/rdma: add inqueue rx chunks perf counter ([pr#14782](#), Haomai Wang)
- rdma: msg/async/rdma: fix log line spacing ([pr#13263](#), Adir Lev)
- rdma: msg/async/rdma: Make poll\_blocking() poll for async events in additio...

([pr#14320](#), Amir Vadai)

- rdma: msg/async/rdma: Make port number an attribute of the Connection not o... ([pr#14297](#), Amir Vadai)
- rdma: msg/async/rdma: RDMA-CM, get\_device() by ibv\_context ([pr#14410](#), Amir Vadai)
- rdma: msg/async: Revert RDMA-CM ([pr#15262](#), Amir Vadai)

Replace using sleep with new wait\_for\_health() bash function

- rgw: abort early when s->length empty during putobj ([pr#15682](#), Jiaying Ren)
- rgw: AbortMultipart request returns NoSuchUpload error if the meta obj doesn't exist ([pr#12793](#), Zhang Shaowen)
- rgw: acl grants num limit ([pr#16291](#), Enming Zhang)
- rgw: add a new error code for non-existed subuser ([pr#16095](#), Zhao Chao)
- rgw: add a new error code for non-existed user ([issue#20468](#), [pr#16033](#), Zhao Chao)
- rgw: add apis to support ragweed ([pr#13645](#), Yehuda Sadeh)
- rgw: add a separate configuration for data notify interval ([pr#16551](#), fang yuxiang)
- rgw: add bucket size limit check to radosgw-admin ([issue#17925](#), [pr#11796](#), Matt Benjamin)
- rgw: Added a globbing method for AWS Policies ([pr#12445](#), Pritha Srivastava)
- rgw: Added code for REST APIs for AWS Roles ([pr#12104](#), Pritha Srivastava)
- rgw: Added code to correctly account for bytes sent/ received during a 'PUT' operation ([pr#14042](#), Pritha Srivastava)
- rgw: Adding code to create tenanted user for s3 bucket policy tests ([pr#15028](#), Pritha Srivastava)
- rgw: add lifecycle validation according to S3 ([issue#18394](#), [pr#12750](#), Zhang Shaowen)
- rgw: add missing RGWPeriod::reflect() based on new atomic update\_latest\_epoch() ([issue#19816](#), [issue#19817](#), [pr#14915](#), Casey Bodley)
- rgw: add -num-zonegroups option for multi test ([pr#14216](#), lvshuhua)
- rgw: add override in header files mostly ([pr#13586](#), liuchang0812)
- rgw: add override in rgw subsystem ([issue#18922](#), [pr#13441](#), liuchang0812)
- rgw: add pool namespace to cache's key so that system obj can have unique key

([issue#19372](#), [pr#14125](#), Zhang Shaowen)

- rgw: add radosclient finisher to perf counter ([issue#19011](#), [pr#13535](#), lvshuhua)
- rgw: add "rgw\_verify\_ssl" config ([pr#15301](#), Shasha Lu)
- rgw: add 'state==SyncState::IncrementalSync' condition when add item ... ([pr#14552](#), Shasha Lu)
- rgw: add support container and object levels of swift bulkupload ([pr#14775](#), Jing Wenjun)
- rgw: add support for delete marker expiration in s3 lifecycle ([issue#19730](#), [pr#14703](#), Zhang Shaowen)
- rgw: add support for FormPost of Swift API ([issue#17273](#), [pr#11179](#), Radoslaw Zarzynski, Orit Wasserman)
- rgw: add support for multipart upload expiration ([issue#19088](#), [pr#13622](#), Zhang Shaowen)
- rgw: add support for noncurrentversion expiration in s3 lifecycle ([issue#18916](#), [pr#13385](#), Zhang Shaowen)
- rgw: add support for Swift's TempURLs with prefix-based scope ([issue#20398](#), [pr#16370](#), Radoslaw Zarzynski)
- rgw: add support for the BulkUpload of Swift API ([pr#12243](#), Radoslaw Zarzynski)
- rgw: add the remove-x-delete feature to cancel swift object expiration ([issue#19074](#), [pr#13621](#), Jing Wenjun)
- rgw: add the Vim's modeline into rgw\_orphan.cc ([pr#15431](#), Radoslaw Zarzynski)
- rgw: add variadic string join for s3 signature generation ([pr#15678](#), Casey Bodley)
- rgw: Add -zonegroup-new-name in usage ([pr#12084](#), Hans van den Bogert)
- rgw: allow larger payload for period commit ([issue#19505](#), [pr#14355](#), Casey Bodley)
- rgw: allow system users to read SLO parts ([issue#19027](#), [pr#13561](#), Casey Bodley)
- rgw: auto reshards old buckets ([pr#15665](#), Orit Wasserman)
- rgw: avoid listing user buckets for rgw\_delete\_user ([pr#13991](#), liuchang0812)
- rgw: avoid using null pointer in rgw\_file.cc ([pr#14474](#), lihongjie)
- rgw: be aware about tenants on cls\_user\_bucket -> rgw\_bucket conversion ([issue#18364](#), [issue#16355](#), [pr#13220](#), Radoslaw Zarzynski)

- rgw: bucket index check in radosgw-admin removes valid index ([issue#18470](#), [pr#12851](#), Zhang Shaowen)
- rgw: bucket stats display bucket index type ([pr#14466](#), fang yuxiang)
- rgw: change default chunk size to 4MB ([issue#18621](#), [issue#18622](#), [issue#18623](#), [pr#13035](#), Yehuda Sadeh)
- rgw: change loglevel to 20 for 'System already converted' message ([issue#18919](#), [pr#13399](#), Vikhyat Umrao)
- rgw: change loglevel to 5 in user's quota sync ([issue#18921](#), [pr#13408](#), Zhang Shaowen)
- rgw: Changes for s3test config file, to add user under a tenant ([pr#15753](#), Pritha Srivastava)
- rgw: check placement existence when create bucket ([pr#16385](#), Jiaying Ren)
- rgw: check placement target existence during bucket creation ([pr#16384](#), Jiaying Ren)
- rgw: civetweb don't go past the array index while calling mg\_start ([issue#19749](#), [pr#14750](#), Abhishek Lekshmanan, Jesse Williamson)
- rgw: clean redundant code ([pr#13302](#), Yankun Li)
- rgw: clean unuse code in cls\_stateglog\_check\_state ([pr#10260](#), weiqiaomiao)
- rgw: clean-up error mapping in Swift's authentication strategy ([pr#15756](#), Radoslaw Zarzynski)
- rgw: cleanup: fix variable name in RGWRados::create\_pool() declaration ([pr#14547](#), Nathan Cutler)
- rgw: cleanup lc continuation ([pr#14906](#), Jiaying Ren)
- rgw: cleanup lifecycle managment ([pr#13820](#), Jiaying Ren)
- rgw: cleanup rgw-admin duplicated judge during OLH GET/READLOG ([pr#15700](#), Jiaying Ren)
- rgw: clean up the redundant assignment in last\_entry\_in\_listing ([pr#13387](#), Jing Wenjun)
- rgw: clean up the unneeded rgw::io::ChunkingFilter::has\_content\_length ([pr#13504](#), Radoslaw Zarzynski)
- rgw: cleanup unused codes in rgw\_admin.cc ([pr#15771](#), fang yuxiang)
- rgw: cleanup unused var in rgw/rgw\_rest\_s3.cc ([pr#13434](#), Jiaying Ren)

- rgw: clear master\_zonegroup when resetting RGWPeriodMap ([issue#17239](#), [pr#12658](#), Orit Wasserman)
- rgw: clear old zone short ids on period update ([issue#15618](#), [pr#13949](#), Casey Bodley)
- rgw: cls: ceph::timespan tag\_timeout wrong units ([issue#20380](#), [pr#16026](#), Matt Benjamin)
- rgw: cls/rgw: Clean up the “magic string” usage in the cls layer for RGW ([pr#12536](#), Ira Cooper)
- rgw: cls/rgw: list\_plain\_entries() stops before bi\_log entries ([issue#19876](#), [pr#14981](#), Casey Bodley)
- rgw: cls/user: cls\_user\_bucket backward compatibility ([issue#19367](#), [pr#14128](#), Yehuda Sadeh)
- rgw: cls\_user don't clobber existing bucket stats when creating bucket ([issue#16357](#), [pr#10121](#), Abhishek Lekshmanan)
- rgw: complete versioning enablement after sending it to meta master ([issue#18003](#), [pr#12444](#), Orit Wasserman)
- rgw: Compress crash bug refactor ([issue#20098](#), [pr#15569](#), Adam Kupczyk)
- rgw: continuation of the auth rework - AWSv4 ([issue#18800](#), [pr#14885](#), Radoslaw Zarzynski, Javier M. Mellid)
- rgw: continuation of the auth rework ([pr#12893](#), Radoslaw Zarzynski, Matt Benjamin)
- rgw: Correcting the condition in ceph\_assert while parsing an AWS Principal ([pr#15997](#), Pritha Srivastava)
- rgw: correct the debug info when unlink instance failed ([pr#13761](#), Zhang Shaowen)
- rgw: Correct the return codes for the health check feature ([issue#19025](#), [pr#13557](#), Pavan Rallabhandi)
- rgw: custom user data header ([issue#19644](#), [pr#14592](#), Pavan Rallabhandi)
- rgw: datalog trim and mdlog trim handles the result returned by osd incorrectly ([issue#20190](#), [pr#15507](#), Zhang Shaowen)
- rgw: data sync includes instance in rgw\_obj\_index\_key ([pr#13948](#), Casey Bodley)
- rgw: deduplicate variants of rgw\_make\_bucket\_entry\_name() ([pr#14299](#), Radoslaw Zarzynski)
- rgw: delete non-empty buckets in slave zonegroup works not well ([issue#19313](#), [pr#14043](#), Zhang Shaowen)

- rgw: delete object in error path ([issue#20620](#), [pr#16324](#), Yehuda Sadeh)
- rgw: disable dynamic reshading for 1st L point release ([pr#16969](#), Matt Benjamin)
- rgw: display more info when using radosgw-admin bucket stats ([pr#15256](#), fang.yuxiang)
- rgw: Do not decrement stats cache when the cache values are zero ([issue#20661](#), [pr#16389](#), Pavan Rallabhandi)
- rgw: Do not fetch bucket stats by default upon bucket listing ([issue#20377](#), [pr#15834](#), Pavan Rallabhandi)
- rgw: do not log debug output at level 0 ([pr#15633](#), Jens Rosenboom)
- rgw: don't do unneccesary write if buffer with zero length ([pr#14925](#), fang yuxiang)
- rgw: don't init rgw\_obj from rgw\_obj\_key when it's incorrect to do so ([issue#19096](#), [pr#13676](#), Yehuda Sadeh)
- rgw: don't log the env\_map twice ([pr#13481](#), Abhishek Lekshmanan)
- rgw: don't read all user input for a few param requests ([pr#13815](#), Abhishek Lekshmanan)
- rgw: don't return skew time error in pre-signed url ([issue#18828](#), [pr#13354](#), liuchang0812)
- rgw: dont spawn error\_repo until lease is acquired ([issue#19446](#), [pr#14714](#), Casey Bodley)
- rgw: don't specify a length when converting bl -> string ([issue#20037](#), [pr#15599](#), Abhishek Lekshmanan)
- rgw: don't use strlen in constexprs to not brake Clang builds ([pr#15688](#), Radoslaw Zarzynski)
- rgw: drop asio/{yield,coroutine}.hpp replacements ([pr#15413](#), Kefu Chai)
- rgw: Drop dump\_usage\_bucket\_info() to silence warning from -Wunused-function ([pr#16497](#), Wei Qiaomiao)
- rgw: drop unused find\_replacement() and some function docs ([pr#16386](#), Jiaying Ren)
- rgw: drop unused function RGWRemoteDataLog::get\_shard\_info() ([pr#16236](#), Shasha Lu)
- rgw: drop unused param "bucket" from select\_bucket\_placement ([pr#14390](#), Shasha Lu)

- rgw: drop unused port var ([pr#14412](#), Jiaying Ren)
- rgw: drop unused rgw\_pool parameter, local variables and member variable ([pr#16154](#), Jiaying Ren)
- rgw: drop unused var header\_ended ([pr#15686](#), Jiaying Ren)
- rgw: drop using std ns in header files and other cleanups ([pr#15137](#), Abhishek Lekshmanan)
- rgw: dynamic resharding ([pr#15493](#), Yehuda Sadeh, Orit Wasserman)
- rgw: enable to update acl of bucket created in slave zonegroup ([issue#16888](#), [pr#14082](#), Guo Zhandong)
- rgw: error\_code in error log is not right when data sync fails ([issue#18437](#), [pr#12810](#), Zhang Shaowen)
- rgw: error more verbosely in RGWRados::create\_pool ([pr#14642](#), Matt Benjamin)
- rgw: external auth engines of S3 honor rgw\_keystone\_implicit\_tenants ([issue#17779](#), [pr#15572](#), Radoslaw Zarzynski)
- rgw: Fix a bug that multipart upload may exceed the quota ([issue#19602](#), [pr#12010](#), Zhang Shaowen)
- rgw: fix asctime when logging in rgw\_lc ([pr#16422](#), Abhishek Lekshmanan)
- rgw: fix break inside of yield in RGWFetchAllMetaCR ([issue#17655](#), [pr#11586](#), Casey Bodley)
- rgw: fix broken /crossdomain.xml, /info and /healthcheck of Swift API ([issue#19520](#), [pr#14373](#), Radoslaw Zarzynski)
- rgw: fix build of conflict after auth rework ([pr#14203](#), Casey Bodley)
- rgw: fix configurable write obj window size ([pr#13934](#), hechuang)
- rgw: fix constexpr for string\_size in clang ([pr#15738](#), Adam C. Emerson, Casey Bodley)
- rgw: fix disabling Swift's object versioning through empty X-Versions-Location ([issue#18852](#), [pr#13303](#), Jing Wenjun)
- rgw: Fix duplicate tag removal during GC ([issue#20107](#), [pr#15912](#), Jens Rosenboom)
- rgw: fix error code of inexistence of versions location in swift api ([issue#18880](#), [pr#13350](#), Jing Wenjun)
- rgw: fix error handling in get\_params() of RGWPostObj\_ObjStore\_S3 ([pr#15670](#), Radoslaw Zarzynski)

- rgw: fix error handling in the link() method of RGWBucket ([issue#20279](#), [pr#15669](#), Radoslaw Zarzynski)
- rgw: fix error message in removing bucket with -bypass-gc flag ([issue#20688](#), [pr#16419](#), Abhishek Varshney)
- rgw: fix err when copy object in bucket with specified placement rule ([issue#20378](#), [pr#15837](#), fang yuxiang)
- rgw: fixes for AWSBrowserUploadAbstractor auth ([issue#20372](#), [pr#15882](#), Radoslaw Zarzynski, Casey Bodley)
- rgw: Fixes typo in rgw\_admin.cc ([issue#19026](#), [pr#13576](#), Ronak Jain)
- rgw: fix for broken yields in RGWMetaSyncShardCR ([issue#18076](#), [pr#12223](#), Casey Bodley)
- rgw: fix for EINVAL errors on forwarded bucket put\_acl requests ([pr#14376](#), Casey Bodley)
- rgw: fix for null version\_id in fetch\_remote\_obj() ([pr#14375](#), Casey Bodley)
- rgw: Fix for Policy Parse exception in case of multiple statements ([pr#16689](#), Pritha Srivastava)
- rgw: fix forward request for bulkupload to be applied in multisite ([issue#19645](#), [pr#14601](#), Jing Wenjun)
- rgw: fix 'gc list -include-all' command infinite loop the first items ([issue#19978](#), [pr#12774](#), Shasha Lu, fang yuxiang)
- rgw: fix get bucket policy s3 compatible issue ([pr#15280](#), Enming Zhang)
- rgw: fix handling of -remote in radosgw-admin period commands ([issue#19554](#), [pr#14407](#), Casey Bodley)
- rgw: fix handling RGWUserInfo::system in RGWHandler\_REST\_SWIFT ([issue#18476](#), [pr#12865](#), Radoslaw Zarzynski)
- rgw: fix infinite loop in rest api for log list ([issue#20386](#), [pr#15983](#), xierui, Casey Bodley)
- rgw: fix init\_bucket\_for\_sync retcode ([pr#13684](#), Shasha Lu)
- rgw: fix lc list failure when shards not be all created ([issue#19898](#), [pr#15025](#), Jiaying Ren)
- rgw: fix leaks with incomplete multipart ([issue#17164](#), [pr#15630](#), Abhishek Varshney)
- rgw: fix marker encoding problem ([issue#20463](#), [pr#15998](#), Marcus Watts)

- rgw: fix memory leak in copy\_obj\_to\_remote\_dest ([pr#9974](#), weiqiaomiao)
- rgw: fix memory leak in delete\_obj\_aio ([pr#13998](#), wangzhengyong)
- rgw: fix memory leak in RGWGetObjLayout ([pr#14014](#), liuchang0812)
- rgw: fix memory leaks during Swift Static Website's error handling ([issue#20757](#), [pr#16531](#), Radoslaw Zarzynski)
- rgw: fix not initialized vars which cause rgw crash with ec data pool ([issue#20542](#), [pr#16177](#), Aleksei Gutikov)
- rgw: fix off-by-one in RGWDataChangesLog::get\_info ([issue#18488](#), [pr#12884](#), Casey Bodley)
- rgw: fix parse/eval of policy conditions with IfExists ([issue#20708](#), [pr#16463](#), Casey Bodley)
- rgw: fix period update crash ([issue#18631](#), [pr#13054](#), Orit Wasserman)
- rgw: fix potential null pointer dereference in rgw\_admin ([pr#15667](#), Radoslaw Zarzynski)
- rgw: fix radosgw-admin data sync run crash ([issue#20423](#), [pr#15938](#), Shasha Lu)
- rgw: fix radosgw-admin retcode ([pr#15257](#), Shasha Lu)
- rgw: fix RadosGW hang during multi-chunk upload of AWSv4 ([issue#19754](#), [pr#14770](#), Radoslaw Zarzynski)
- rgw: fix radosgw will crash when service is restarted during lifecycl... ([issue#20756](#), [pr#16495](#), Wei Qiaomiao)
- rgw: fix response header of Swift API ([issue#19443](#), [pr#14280](#), tone-zhang)
- rgw: fix rest client's order of args in get\_v2\_signature ([pr#15731](#), Casey Bodley)
- rgw: fix rgw bucket policy IfExists position ([issue#20248](#), [pr#15607](#), yuliyang)
- rgw: fix rgw hang when do RGWRealmReloader::reload after go SIGHUP ([issue#20686](#), [pr#16417](#), fang.yuxiang)
- rgw: fix RGWPutBucketPolicy error when set BucketPolicy again without delete pre set Policy ([issue#20252](#), [pr#15617](#), yuliyang)
- rgw: fix s3 object uploads with chunked transfers and v4 signatures ([issue#20447](#), [pr#15965](#), Marcus Watts)
- rgw: fix segfault in RevokeThread during its shutdown procedure ([issue#19831](#), [pr#15033](#), Radoslaw Zarzynski)
- rgw: fix slave zonegroup cannot enable the bucket versioning ([issue#18003](#),

- pr#12175, lvshuhua)
- rgw: fix SLO/DLO range requests (pr#15060, Shasha Lu)  
rgw: fix swift cannot disable object versioning
  - rgw: fix swift default auth error after auth strategy refactoring (pr#15711, Casey Bodley)
  - rgw: fix test\_multi.py default config file path (pr#15306, Jiaying Ren)
  - rgw: fix the bug that part's index can't be removed after completing multipart upload (issue#19604, pr#14500, Zhang Shaowen)
  - rgw: fix the signature mismatch of FormPost in swift API (issue#20220, pr#15564, Jing Wenjun)
  - rgw: fix the UTF8 check on bucket entry name in rgw\_log\_op() (issue#20779, pr#16604, Radoslaw Zarzynski)
  - rgw: fix transition from full to incremental meta sync (pr#13920, Casey Bodley)
  - rgw: fix typo in comment (pr#13578, liuchang0812)
  - rgw: fix uninitialized fields (pr#14120, wangzhengyong)
  - rgw: fix upgrade from hammer when zone doesn't have zoneparams (issue#19231, pr#13900, Orit Wasserman)
  - rgw: Fix up to 1000 entries at a time in check\_bad\_index\_multipart (issue#20772, pr#16692, Orit Wasserman)
  - rgw: fix use of marker in List::list\_objects() (issue#18331, pr#13147, Yehuda Sadeh)
  - rgw: fix versioned bucket data sync fail when upload is busy (issue#18208, pr#12357, lvshuhua)
  - rgw: fix wrong error code for expired Swift TempURL's links (issue#20384, pr#15850, Radoslaw Zarzynski)
  - rgw: fix X-Object-Meta-Static-Large-Object in SLO download (issue#19951, pr#15045, Shasha Lu)
  - rgw: fix zone didn't update realm\_id when added to zonegroup (issue#17995, pr#12139, Tianshan Qu)
  - rgw: forward RGWPutBucketPolicy to meta master (issue#20297, pr#15736, Casey Bodley)
  - rgw: get torrent request's parameter is not the same as amazon s3 (issue#19136, pr#13760, Zhang Shaowen)

- rgw: get wrong content when download object with specific range with compression ([issue#20100](#), [pr#15323](#), fang yuxiang)
- rgw: handle error return value in build\_linked\_oids\_index ([pr#13955](#), wangzhengyong)
- rgw: http\_client clarify the debug msg function call ([pr#13688](#), Abhishek Lekshmanan)
- rgw: if user.email is empty, dont try to delete ([issue#18980](#), [pr#13783](#), Casey Bodley)
- rgw: implement get/put object tags for S3 ([pr#13753](#), Abhishek Lekshmanan)
- rgw: improve handling of illformed Swift's container ACLs ([issue#18796](#), [pr#13248](#), Radoslaw Zarzynski)
- rgw: /info claims we do support Swift's accounts ACLs ([issue#20394](#), [pr#15887](#), Radoslaw Zarzynski)
- rgw: initialize non-static class members in ESQueryCompiler ([pr#15884](#), Jos Collin)
- rgw: initialize Non-static class member val in ESQueryNodeLeafVal\_Int ([pr#15888](#), Jos Collin)
- rgw: initialize Non-static class member worker in RGWReshard ([pr#15886](#), Jos Collin)
- rgw: Initialize of member variable admin\_specified in RGWUserAdminOpState ([pr#16847](#), amitkuma)
- rgw: Initialize pointer fields ([pr#16021](#), Jos Collin)
- rgw: LCWorker's worktime is not the same as config rgw\_lifecycle\_work\_time ([issue#18087](#), [pr#11963](#), Zhang Shaowen)
- rgw: ldap: simple\_bind() should set ldap version option on tldap ([pr#12616](#), Weibing Zhang)
- rgw: lease\_stack: use reset method instead of assignment ([pr#16185](#), Nathan Cutler)
- rgw: Let the object stat command be shown in the usage ([issue#19013](#), [pr#13291](#), Pavan Rallabhandi)
- rgw: librgw shut ([issue#18585](#), [pr#12972](#), Matt Benjamin)
- rgw: lifecycle thread shouldn't process the bucket which has been deleted ([issue#20285](#), [pr#15677](#), Zhang Shaowen)
- rgw: lock is not released when set sync marker is failed ([issue#18077](#), [pr#12197](#),

Zhang Shaowen)

- rgw: log\_meta only for more than one zone ([issue#20357](#), [pr#15777](#), Orit Wasserman, Leo Zhang)
- rgw: lower some log's level in gc process ([pr#15426](#), Zhang Shaowen)
- rgw: luminous: rgw: Fix rgw not responding occasionally when receiving SIGHUP signal ([issue#20962](#), [pr#17113](#), Yao Zongyou)
- rgw: luminous: RGW: Get Bucket ACL does not honor the s3:GetBucketACL action ([issue#21013](#), [issue#21056](#), [pr#17117](#), Abhishek Lekshmanan)
- rgw: luminous: rgw: GetObject Tagging needs to exit earlier if the object has no attributes ([issue#21054](#), [issue#21010](#), [pr#17118](#), Abhishek Lekshmanan)
- rgw: luminous: rgw\_lc: support for AWSv4 authentication ([pr#16914](#), Abhishek Lekshmanan)
- rgw: luminous: rgw: S3 v4 auth fails when query string contains ([issue#21000](#), [issue#21003](#), [issue#20501](#), [issue#21043](#), [pr#17114](#), Zhang Shaowen, Marcus Watts)
- rgw: luminous: rgw: Use namespace for lc\_pool and roles\_pool ([issue#20177](#), [issue#20967](#), [pr#16943](#), Orit Wasserman)
- rgw: make RGWEnv return a const ref. to its map ([pr#15269](#), Abhishek Lekshmanan)
- rgw: make sending Content-Length in 204 and 304 responses controllable ([issue#16602](#), [pr#10156](#), Radoslaw Zarzynski)
- rgw: make sync thread name clear ([issue#18860](#), [pr#13324](#), lvshuhua)
- rgw: match wildcards in StringLike policy conditions ([issue#20308](#), [pr#16491](#), Casey Bodley)
- rgw: metadata search part 2 ([pr#14351](#), Yehuda Sadeh)
- rgw: meta sync thread crash at RGWMetaSyncShardCR ([issue#20251](#), [pr#15660](#), fang.yuxiang)
- rgw: migrate atomic\_t to std::atomic<> (ebirah) ([pr#14839](#), Jesse Williamson)
- rgw: migrate atomic\_t to std::atomic<> ([pr#15001](#), Jesse Williamson)
- rgw: modify email to empty by admin RESTful api doesn't work ([pr#16309](#), fang.yuxiang)
- rgw: move the S3 anonymous auth handling to a dedicated engine ([pr#16485](#), Radoslaw Zarzynski)
- rgw: multipart copy-part remove '/' for s3 java sdk request header ([issue#20075](#), [pr#15283](#), root)

- rgw: multisite enabled over multiple clusters ([pr#12535](#), Ali Maredia)
- rgw: multisite: fixes for zonegroup redirect ([issue#19488](#), [pr#14319](#), Casey Bodley)
- rgw:multisite: fix RGWRadosRemoveOmapKeysCR and change cn to intrusive\_ptr ([issue#20539](#), [pr#16197](#), Shasha Lu)
- rgw: never let http\_redirect\_code of RGWRedirectInfo to stay uninitialized ([issue#20774](#), [pr#16601](#), Radoslaw Zarzynski)
- rgw: omit X-Account-Access-Control if there is no grant to serialize ([issue#20395](#), [pr#15883](#), Radoslaw Zarzynski)
- rgw: only log metadata on metadata master zone ([issue#20244](#), [pr#15613](#), Casey Bodley)
- rgw: optimize data sync. Add zones\_trace in log to avoid needless sync ([issue#19219](#), [pr#13851](#), Zhang Shaowen)
- rgw: optimize generating torrent file. Object data won't stay in memory now ([pr#15153](#), Zhang Shaowen)
- rgw: orphan: fix error messages ([pr#12782](#), Weibing Zhang)
- rgw: pass authentication domain to civetweb ([issue#17657](#), [pr#12861](#), Abhishek Lekshmanan)
- rgw: polymorphic error codes ([pr#10690](#), Pritha Srivastava, Marcus Watts)
- rgw: print is\_admin as int instead of \_\_u8 ([pr#12264](#), Casey Bodley)
- rgw: put object's acl can't work well on the latest object ([issue#18649](#), [pr#13078](#), Zhang Shaowen)
- rgw: radosgw-admin: use zone id when creating a zone ([issue#19498](#), [pr#14340](#), Orit Wasserman)
- rgw: radosgw-admin: warn that 'realm rename' does not update other clusters ([issue#19746](#), [pr#14722](#), Casey Bodley)
- rgw: radosgw, crypto: simplified code in handle\_data functions ([pr#15598](#), Adam Kupczyk)
- rgw: radosgw: fix compilation with cryptopp ([pr#15960](#), Adam Kupczyk)
- rgw: raise debug level of meta sync logging ([pr#15524](#), Casey Bodley)
- rgw: raise debug level of RGWPostObj\_ObjStore\_S3::get\_policy ([pr#16203](#), Shasha Lu)
- rgw: reject request if decoded URI contains 0 in the middle ([issue#20418](#),

pr#15953, Radoslaw Zarzynski)

- rgw: remove a redundant judgement in rgw\_rados.cc:delete\_obj (pr#11124, Zhang Shaowen)
- rgw: Removed Unwanted headers (pr#14183, Jos Collin)
- rgw: remove duplicate flush formatter (pr#12437, Guo Zhandong)
- rgw: remove extra RGWMPObj in rgw\_multi.h (pr#14619, Casey Bodley)
- rgw: remove fastcgi from default rgw frontends (pr#15098, Casey Bodley)
- rgw: remove invalid read size4 (issue#18071, pr#12767, Matt Benjamin)  
rgw: Remove pessimizing move
- rgw: remove redundant codes in rgw\_cache.h (pr#13902, lihongjie)
- rgw: Remove spurious XML header for GetBucketPolicy (issue#20247, pr#15586, Adam C. Emerson)
- rgw: remove the useless output when listing zonegroups (pr#16331, Zhang Shaowen)
- rgw: remove unused func in rgw\_file.h (pr#15698, lihongjie)
- rgw: remove useless -tier\_type in radosgw-admin (pr#13856, Zhang Shaowen)
- rgw: rename s3\_code to err\_code for swift (pr#12300, Guo Zhandong)
- rgw: Replace get\_zonegroup().is\_master\_zonegroup() with is\_meta\_master() in RGWBulkDelete::Deleter::delete\_single() (pr#16062, Fan Yang)
- rgw: req xml params size limitation error msg (pr#16310, Enming Zhang)
- rgw: respect Swift's negative, HTTP referer-based ACL grants (issue#18841, pr#14344, Radoslaw Zarzynski)
- rgw: restore admin socket path in mrgw.sh (pr#16540, Casey Bodley)
- rgw: return the version id in get object and object metadata request (issue#19370, pr#14117, Zhang Shaowen)
- rgw: rgw-admin: fix bucket limit check argparse, div(0) (pr#15316, Matt Benjamin)
- rgw: rgw-admin: remove deprecated regionmap commands (issue#18725, pr#13963, Casey Bodley)
- rgw: rgw\_common: use string::npos for the results of str.find (pr#14341, Abhishek Lekshmanan)
- rgw: rgw\_crypt: log error messages during failures (pr#16726, Abhishek Lekshmanan)

- rgw: rgw\_file: add compression interop to RGW NFS ([issue#20462](#), [pr#15989](#), Matt Benjamin)
- rgw: rgw\_file: add lock protection for readdir against gc ([issue#20121](#), [pr#15329](#), Gui Hecheng)
- rgw: rgw\_file: add service map registration ([pr#16251](#), Matt Benjamin)
- rgw: rgw\_file: add timed namespace invalidation ([issue#18651](#), [pr#13038](#), Matt Benjamin)
- rgw: rgw\_file: avoid a recursive lane lock in LRU drain ([issue#20374](#), [pr#15819](#), Matt Benjamin)
- rgw: rgw\_file: avoid stranding invalid-name bucket handles in fhcache ([issue#19036](#), [pr#13590](#), Matt Benjamin)
- rgw: rgw\_file cleanup names ([pr#15568](#), Gui Hecheng)
- rgw: rgw\_file: cleanup virtual keyword on derived functions ([pr#14908](#), Gui Hecheng)
- rgw: rgw\_file: ensure valid\_s3\_object\_name for directories, too ([issue#19066](#), [pr#13614](#), Matt Benjamin)
- rgw: rgw\_file: fix assert upon setattr on bucket ([issue#20287](#), [pr#15679](#), Gui Hecheng)
- rgw: rgw\_file: fix double unref on rgw\_fh for rename ([pr#13988](#), Gui Hecheng)
- rgw: rgw\_file: fix flags set on unsuccessful unlink ([pr#15222](#), Gui Hecheng)
- rgw: rgw\_file: fix fs\_inst progression ([issue#19214](#), [pr#13832](#), Matt Benjamin)
- rgw: rgw\_file: fix missing unlock in unlink ([issue#19435](#), [pr#14262](#), Gui Hecheng)
- rgw: rgw\_file: fix misuse of make\_key\_name before make\_fhk ([pr#15108](#), Gui Hecheng)
- rgw: rgw\_file: fix non-negative return code for open operation ([pr#14045](#), Gui Hecheng)
- rgw: rgw\_file: fix non-posix errcode EINVAL to ENAMETOOLONG ([pr#13764](#), Gui Hecheng)
- rgw: rgw\_file: fix readdir after dirent-change ([issue#19634](#), [pr#14561](#), Matt Benjamin)
- rgw: rgw\_file: fix reversed return value of getattr ([pr#13895](#), Gui Hecheng)
- rgw: rgw\_file: fix RGWLibFS::setattr for directory objects ([issue#18808](#), [pr#13252](#), Matt Benjamin)

- rgw: rgw\_file: fix up potential race condition ([pr#14553](#), Gui Hecheng)
- rgw: rgw\_file: implement reliable has-children check (unlink dir) ([issue#19270](#), [pr#13953](#), Matt Benjamin)
- rgw: rgw\_file: interned RGWFileHandle objects need parent refs ([issue#18650](#), [pr#13084](#), Matt Benjamin)
- rgw: rgw\_file: posix style atime,ctime,mtime ([pr#13765](#), Gui Hecheng)
- rgw: rgw\_file: pre-compute unix attrs in write\_finish() ([issue#19653](#), [pr#14609](#), Matt Benjamin)
- rgw: rgw\_file: prevent conflict of mkdir between restarts ([issue#20275](#), [pr#15655](#), Gui Hecheng)
- rgw: rgw\_file: properly & or flags ([issue#20663](#), [pr#16448](#), Matt Benjamin)
- rgw: rgw\_file: release rgw\_fh lock and ref on ENOTEMPTY ([issue#20061](#), [pr#15246](#), Matt Benjamin)
- rgw: rgw\_file: removed extra rele() on fs in rgw\_umount() ([pr#15152](#), Gui Hecheng)
- rgw: rgw\_file: remove hidden uxattr objects from buckets on delete ([issue#20045](#), [pr#15210](#), Matt Benjamin)
- rgw: rgw\_file: remove post-unlink lookup check ([issue#20047](#), [pr#15216](#), Matt Benjamin)
- rgw: rgw\_file: replace raw fs->fh\_lru.unref with predefined fs->unref ([pr#15541](#), Gui Hecheng)
- rgw: rgw\_file: RGWFileHandle dtor must also cond-unlink from FHCache ([issue#19112](#), [pr#13712](#), Matt Benjamin)
- rgw: rgw\_file skip policy read for virtual components ([pr#16034](#), Gui Hecheng)
- rgw: rgw\_file: split last argv on ws, if provided ([pr#12965](#), Matt Benjamin)
- rgw: rgw\_file: store bucket uxattrs on the bucket ([issue#20082](#), [pr#15293](#), Matt Benjamin)
- rgw: rgw\_file: support readdir cb type hints (plus fixes) ([issue#19623](#), [issue#19625](#), [issue#19624](#), [pr#14458](#), Matt Benjamin)
- rgw: rgw\_file: use fh\_hook::is\_linked() to check residence ([issue#19111](#), [pr#13703](#), Matt Benjamin)
- rgw: rgw\_file: v3: fix write-timer action ([issue#19932](#), [pr#15097](#), Matt Benjamin)
- rgw: rgw : fix race in RGWCompleteMultipart ([issue#20861](#), [pr#16732](#), Abhishek Varshney)

- rgw: rgw: fix s3 aws v2 signature priority between header['X-Amz-Date'] and header['Date'] ([issue#20176](#), [pr#15467](#), yuliyang)
- rgw: rgw: fix the subdir without slash of s3 website url ([issue#20307](#), [pr#15703](#), liuhong)
- rgw: rgw\_lc: drop a bunch of unused headers ([pr#14342](#), Abhishek Lekshmanan)
- rgw: rgw\_ldap: log the ldap err in case of bind failure ([pr#14781](#), Abhishek Lekshmanan)
- rgw: rgw/lifecycle: do not send lifecycle rules when GetLifeCycle failed ([issue#19363](#), [pr#14160](#), liuchang0812)
- rgw: RGWMetaSyncShardControlCR retries with backoff on all error codes ([issue#19019](#), [pr#13546](#), Casey Bodley)
- rgw: RGWMetaSyncShardCR drops stack refs on destruction ([issue#18412](#), [issue#18300](#), [pr#12605](#), Casey Bodley)
- rgw: rgw multisite: automated mdlog trimming ([pr#13111](#), Casey Bodley)
- rgw: rgw multisite: feature of bucket sync enable/disable ([pr#15801](#), Zhang Shaowen, Casey Bodley, Zengran Zhang)
- rgw: rgw multisite: fixes for meta sync across periods ([issue#18639](#), [pr#13070](#), Casey Bodley)
- rgw: rgw multisite: fix ref counting of completions ([issue#18414](#), [issue#18407](#), [pr#12841](#), Casey Bodley)
- rgw: rgw-multisite: fix the problem of rgw website configure 'RedirectAllRequestsTo' failed to sync to slave zone ([pr#15036](#), yuliyang)
- rgw: rgw-multisite: fix the problem of rgw website configure request not redirect to metadata master ([pr#15082](#), yuliyang)
- rgw: rgw multisite: remove the redundant post in OPT\_ZONEGROUP MODIFY ([pr#14359](#), Jing Wenjun)
- rgw: rgw/multisite: validate bucket location during bucket creation ([pr#15333](#), Jiaying Ren)
- rgw: RGW NFS: add nfs.rst to doc/radosgw ([pr#15789](#), Matt Benjamin)
- rgw: rgw\_op: remove unused variable iter ([pr#14276](#), Weibing Zhang)
- rgw: RGWPeriodPusher spawns http thread before cr thread ([issue#19834](#), [pr#14936](#), Casey Bodley)
- rgw: rgw\_rados: create sync module instances only if run\_sync\_thread is set ([issue#19830](#), [pr#14994](#), Abhishek Lekshmanan)

- rgw: rgw\_rados drop deprecated global var ([pr#14411](#), Jiaying Ren)
- rgw: rgw\_rados: initialize cur\_shard ([pr#15735](#), Abhishek Lekshmanan)
- rgw: rgw realm set fixes ([issue#18333](#), [pr#12731](#), Orit Wasserman)
- rgw: rgw/rgw\_frontend.h: Return negative value for empty uid in RGWLoadGenFrontend::init() ([pr#16204](#), jimifm)
- rgw: rgw/rgw\_main.cc: fix parenteses and function result ([pr#12295](#), Willem Jan Withagen)
- rgw: rgw/rgw\_op:Prevents memory leaks when calling func swift\_versioning\_copy() fails ([pr#15328](#), jimifm)
- rgw: rgw/rgw\_rados: Remove duplicate calls in RGWRados::finalize() ([pr#15281](#), jimifm)
- rgw: rgw/rgw\_string.h: FreeBSD would like errno.h included ([pr#15737](#), Willem Jan Withagen)
- rgw: rgw/rgw\_swift\_auth.cc: using string::back() instead as the C++11 recommend ([pr#14827](#), liuyuhong)
- rgw: rgw structures rework ([issue#17996](#), [issue#19249](#), [pr#11485](#), Yehuda Sadeh)
- rgw: rgw,test: fix rgw placement rule pool config option ([pr#16084](#), Jiaying Ren)
- rgw: S3 lifecycle now supports expiration date ([pr#15807](#), Zhang Shaowen)
- rgw: s3 server-side encryption (SSE-C, SSE-KMS) ([pr#11049](#), Adam Kupczyk, Casey Bodley, Radoslaw Zarzynski)
- rgw: segment fault when shard id out of range ([issue#19732](#), [pr#14389](#), redickwang)
- rgw: set dumpable flag after setuid post ff0e521 ([issue#19089](#), [pr#13657](#), Brad Hubbard)
- rgw: set FCGI\_INCLUDE\_DIR for cephd\_rgw\_base ([issue#18918](#), [pr#13393](#), David Disseldorp)
- rgw: set object accounted size correctly ([issue#20071](#), [pr#14950](#), fang yuxiang)
- rgw: set placement rule properly ([pr#15221](#), fang.yuxiang)
- rgw: should delete in\_stream\_req if conn->get\_obj(...) return not zero value ([pr#9950](#), weiqiaomiao)
- rgw: should not restrict location\_constraint same when user not provide ([pr#16770](#), Tianshan Qu)
- rgw: should unlock when reshard\_log->update() reture non-zero in RGWB... ([pr#16502](#),

Wei Qiaomiao)

- rgw: silence compile warning from -Wmaybe-uninitialized ([pr#15996](#), Jiaying Ren)
- rgw: silence warning from -Wmaybe-uninitialized ([pr#15949](#), Jos Collin)
- rgw: stat requests skip compression, manifest handling, etc ([pr#14109](#), Casey Bodley)
- rgw: Support certain archaic and antiquated distributions([pr#15498](#), Adam C. Emerson)
- rgw: swift: ability to update swift read and write acls separately ([issue#19289](#), [pr#14499](#), Marcus Watts)
- rgw: swift: disable revocation thread if sleep == 0 ([issue#19499](#), [issue#9493](#), [pr#14501](#), Marcus Watts)
- rgw: swift: fix anonymous user's error code of getting object ([issue#18806](#), [pr#13242](#), Jing Wenjun)
- rgw: swift: the http referer acl in swift API should be shown ([issue#18665](#), [pr#13003](#), Jing Wenjun)
- rgw: swift: The http referer should be parsed to compare in swift API ([issue#18685](#), [pr#13005](#), Jing Wenjun)
- rgw: switch from "timegm()" to "internal\_timegm()" for better portability ([issue#12863](#), [pr#14327](#), Rishabh Kumar)
- rgw: switch to std::array in RGWBulkUploadOp due to C++11 and FreeBSD ([pr#14314](#), Radoslaw Zarzynski)
- rgw: sync status compares the current master period ([issue#18064](#), [pr#12907](#), Abhishek Lekshmanan)
- rgw: test,rgw: fix rgw placement rule pool config option ([pr#16380](#), Jiaying Ren)
- rgw,tests: luminous: qa/rgw: use 'ceph osd pool application enable' on created pools ([pr#17259](#), Casey Bodley)
- rgw,tests: qa/rgw: add cluster name to path when s3tests scans rgw log ([pr#14845](#), Casey Bodley)
- rgw,tests: qa/rgw: add configuration for server-side encryption tests ([pr#13597](#), Casey Bodley)
- rgw,tests: qa/rgw: add encryption config for s3tests under thrash ([pr#15694](#), Casey Bodley)
- rgw,tests: qa/rgw: add multisite suite to configure and run multisite tests ([pr#14688](#), Casey Bodley)

- rgw,tests: qa/rgw: disable lifecycle tests because of expiration failures ([pr#16760](#), Casey Bodley)
- rgw,tests: qa/rgw: don't scan radosgw logs for encryption keys on jewel upgrade test ([pr#14697](#), Casey Bodley)
- rgw,tests: qa/rgw: fix assertions in radosgw\_admin task ([pr#14842](#), Casey Bodley)
- rgw,tests: qa/rgw: remove apache/fastcgi and radosgw-agent tests ([pr#15184](#), Casey Bodley)
- rgw,tests: qa/suites/rgw/thrash: add osd thrashing tests ([pr#13445](#), Sage Weil)
- rgw,tests: qa/tasks: S3A hadoop task to test s3a with Ceph ([pr#14624](#), Vasu Kulkarni)
- rgw,tests: test/rgw: add bucket acl and versioning tests to test\_multi.py ([pr#12449](#), Casey Bodley)
- rgw,tests: test/rgw: add test for versioned object sync ([pr#12474](#), Casey Bodley)
- rgw,tests: test/rgw: fixes for test\_multi\_period\_incremental\_sync() ([pr#13067](#), Casey Bodley)
- rgw,tests: test/rgw: fix for empty lists as default arguments ([pr#14816](#), Casey Bodley)
- rgw,tests: test/rgw: refactor test\_multi.py for use in qa suite ([pr#14433](#), Casey Bodley)
- rgw,tests: test/rgw: test\_bucket\_delete\_notempty in test\_multi.py ([pr#14090](#), Casey Bodley)
- rgw,tests: vstart: add rgw configuration needed to pass all s3tests ([pr#15782](#), Casey Bodley)
- rgw,tests: vstart: remove rgw\_enable\_static\_website ([pr#15856](#), Casey Bodley)
- rgw: the swift container acl should support field .ref ([issue#18484](#), [pr#12874](#), Jing Wenjun)
- rgw: Turn off fcgi as a frontend ([issue#16784](#), [pr#15070](#), Thomas Serlin)
- rgw: Uninitialized member in LCRule ([pr#15827](#), Jos Collin)
- rgw: update Beast for streaming reads in asio frontend ([pr#14273](#), Casey Bodley)
- rgw: update bucket cors in secondary zonegroup should forward to master ([issue#16888](#), [pr#15260](#), Shasha Lu)
- rgw: update function doc in rgw\_rados.h and rgw\_rados.cc ([pr#15803](#), Jiaying Ren)

- rgw: update is\_truncated in function rgw\_read\_user\_buckets ([issue#19365](#), [pr#14343](#), liuchang0812)
- rgw: usage ([issue#16191](#), [pr#14287](#), Ji Chen, Orit Wasserman)
- rgw: use 64-bit offsets for compression ([issue#20231](#), [pr#15656](#), Adam Kupczyk, fang yuxiang)
- rgw: use a namespace for rgw reshards pool for upgrades as well ([issue#20289](#), [pr#16368](#), Karol Mroz, Abhishek Lekshmanan)
- rgw: Use comparison instead of assignment ([pr#16653](#), amitkuma)
- rgw: Use decoded URI when verifying TempURL ([issue#18590](#), [pr#13007](#), Michal Koutný)
- rgw: use get\_data\_extra\_pool() when get extra pool ([issue#20064](#), [pr#15219](#), fang yuxiang)
- rgw: use pre-defined calls to replace raw flag operation ([pr#15107](#), Gui Hecheng)
- rgw: use rgw\_zone\_root\_pool for region\_map like is done in hammer ([issue#19195](#), [pr#13928](#), Orit Wasserman)
- rgw: use separate http\_manager for read\_sync\_status ([issue#19236](#), [pr#13660](#), Shasha Lu)
- rgw: use uncompressed size for range\_to\_ofs() in slo/dlo ([pr#15931](#), Casey Bodley)
- rgw: using RGW\_OBJ\_NS\_MULTIPART in check\_bad\_index\_multipart ([pr#15774](#), Shasha Lu)
- rgw: using the same bucket num\_shards as master zg when create bucket in secondary zg ([issue#19745](#), [pr#14388](#), Shasha Lu)
- rgw: validate tenant names during user create ([pr#16442](#), Abhishek Lekshmanan)
- rgw: verify md5 in post obj ([issue#19739](#), [pr#14961](#), Yehuda Sadeh)
- rgw: version id doesn't work in fetch\_remote\_obj ([pr#14010](#), Zhang Shaowen)
- rgw: VersionIdMarker and NextVersionIdMarker should be returned when listing object versions ([issue#19886](#), [pr#15014](#), Zhang Shaowen)
- rgw: warning, output may be truncated before the last format character ([pr#14194](#), Jos Collin)
- rgw: when create\_bucket use the same num\_shards with info.num\_shards ([issue#19745](#), [pr#15010](#), Shasha Lu)
- rgw: wip dir orphan ([issue#18992](#), [issue#18989](#), [issue#19018](#), [issue#18991](#), [pr#13529](#), Matt Benjamin)

- rgw: Wip librgw refcnt ([pr#13405](#), Matt Benjamin)
- rgw: wip parentref ([issue#19060](#), [issue#19059](#), [pr#13607](#), Matt Benjamin)
- rgw: Wip rgw fix prefix list ([issue#19432](#), [pr#15916](#), Giovani Rinaldi, Orit Wasserman)
- rgw: Wip rgw openssl 7 ([issue#11239](#), [issue#16535](#), [pr#11776](#), Yehuda Sadeh, Marcus Watts)
- rgw: wip: rgw: rest\_admin/user avoid double checking input args ([pr#13460](#), Abhishek Lekshmanan)
- tests: Add integration tests for admin socket output ([pr#15223](#), Brad Hubbard)
- tests: add MGR=1 so 'pg dump' won't be blocked ([pr#14266](#), Kefu Chai)
- tests: Add openstack requirements to smoke suite ([pr#12913](#), Zack Cerza)
- tests: add setup/teardown for asok dir ([pr#16523](#), Kefu Chai)
- tests: buildpackages: remove because it does not belong ([issue#18846](#), [pr#13297](#), Loic Dachary)
- tests: ceph-disk: add setting for external py-modules for tox-testing ([pr#15433](#), Willem Jan Withagen)
- tests: ceph-disk: use communicate() instead of wait() for output ([pr#16347](#), Kefu Chai)
- tests: ceph-helpers.sh reduce get\_timeout\_delays() verbosity ([pr#13257](#), Kefu Chai)
- tests: ceph\_objectstore\_tool.py: kill all daemons ([pr#14428](#), Kefu Chai)
- tests: ceph\_test\_objectstore: tolerate fsck EOPNOTSUPP too ([pr#13325](#), Sage Weil)
- tests: ceph\_test\_rados\_api\_tier: tolerate ENOENT from 'pg scrub' ([pr#14807](#), Sage Weil)
- tests: ceph\_test\_rados\_api\_watch\_notify: move global variables into test class ([issue#18395](#), [pr#12751](#), Kefu Chai)
- tests: ceph\_test\_rados\_api\_watch\_notify: test timeout using rados\_wat... ([issue#19312](#), [pr#14061](#), Kefu Chai)
- tests: cephtool/test.sh error on full tests ([issue#19698](#), [pr#14647](#), Willem Jan Withagen, David Zafman)
- tests: cephtool/test.sh: Only delete a test pool when no longer needed ([pr#16443](#), Willem Jan Withagen)

- tests: cls\_lock: move lock\_info\_t definition to cls\_lock\_types.h ([pr#16091](#), runsiisi)
- tests: config\_opts: drop unused opts ([pr#15031](#), Kefu Chai)
- tests: Decreased amount of jobs on master, kraken, luminous runs ([pr#17074](#), Yuri Weinstein)
- tests: Don't dump core when using EXPECT\_DEATH ([pr#14821](#), Kefu Chai, Brad Hubbard)
- tests: drop buildpackages.py ([issue#18846](#), [pr#13319](#), Nathan Cutler)
- tests: drop obsolete Perl scripts ([pr#13951](#), Nathan Cutler)
- tests: drop rbd\_cli\_tests.pl and RbdLib.pm ([issue#14825](#), [pr#12821](#), Nathan Cutler)
- tests: drop unused rbd\_functional\_tests.pl script ([issue#14825](#), [pr#12818](#), Nathan Cutler)
- tests: fio\_ceph\_objectstore: fixes improper write request data lifetime ([pr#14338](#), Adam Kupczyk)
- tests: fix broken links in upgrade/hammer-jewel-x/stress-split ([issue#19793](#), [pr#14831](#), Nathan Cutler)
- tests: fix NULL references to be acceptable by Clang ([pr#12880](#), Willem Jan Withagen)
- tests: fix rados/upgrade/jewel-x-singleton and make workunit task handle repo URLs not ending in ".git" ([issue#20554](#), [issue#20368](#), [pr#16228](#), Nathan Cutler, Sage Weil)
- tests: fix regression in qa/tasks/ceph\_master.py ([issue#16263](#), [pr#13279](#), Nathan Cutler, Kefu Chai)
- tests: fix template specialization of PromoteRequest class ([pr#12815](#), Ricardo Dias)
- tests: ignore bogus ceph-objectstore-tool error in ceph\_manager ([issue#16263](#), [pr#13194](#), Nathan Cutler)
- tests: include/denc: support ENCODE\_DUMP ([pr#14962](#), Sage Weil)
- tests: libradosstriper: do not assign garbage to returned value ([pr#15009](#), Kefu Chai)
- tests: luminous: tests: qa/standalone: misc fixes ([issue#20465](#), [issue#20921](#), [issue#20979](#), [pr#16985](#), David Zafman)
- tests: mgr,os,test: kill clang analyzer warnings ([pr#16227](#), Kefu Chai)

- tests: move swift.py task from teuthology to ceph, phase one (master) ([issue#20392](#), [pr#15859](#), Nathan Cutler, Sage Weil, Warren Usui, Greg Farnum, Ali Maredia, Tommi Virtanen, Zack Cerza, Sam Lang, Yehuda Sadeh, Joe Buck, Josh Durgin)
- tests: nosetests: use /usr/bin/env to find nosetests ([pr#12091](#), Willem Jan Withagen)
- tests: os: Argument cannot be negative ([pr#16688](#), amitkuma)
- tests: os/bluestore,test/ceph\_test\_objectstore: silence gcc warnings ([pr#13924](#), Kefu Chai)
- tests: osd-scrub-repair.sh disable scrub backoff in test ([pr#13334](#), Kefu Chai)
- tests: qa: Added luminous to the mix in schedule\_subset.sh ([pr#16430](#), Yuri Weinstein)
- tests: qa/added overrides ([pr#14917](#), Yuri Weinstein)
- tests: qa: Add reboot case for systemd test ([issue#19717](#), [pr#14229](#), Vasu Kulkarni)
- tests: qa: add supported distros for ceph-ansible ([pr#13711](#), Tamil Muthamizhan)
- tests: qa: add task for dnsmasq configuration ([pr#15071](#), Casey Bodley)
- tests: [qa/ceph-deploy]: run create mgr nodes as well ([pr#16216](#), Vasu Kulkarni)
- tests: qa: Cleaned up distros to use latest versions ([pr#12804](#), Yuri Weinstein)
- tests: qa/distros: make centos\_latest 7.3 ([pr#12944](#), Sage Weil)
- tests: qa,doc: document and fix tests for pool application warnings ([pr#16568](#), Sage Weil)
- tests: qa: do not mention ceph branch explicitly ([pr#13225](#), Tamil Muthamizhan)
- tests: qa: do not restrict valgrind runs to centos ([issue#18126](#), [pr#15893](#), Greg Farnum)
- tests: qa/erasure-code: override min\_size to 2 ([issue#19770](#), [pr#14872](#), Kefu Chai)
- tests: qa: fixed distros links ([pr#12770](#), Yuri Weinstein)
- tests: qa/run-standalone.sh: fix the find option to be compatible with GNU find ([pr#16646](#), Kefu Chai)
- tests: qa: specify client for fs workunit ([pr#12914](#), Tamil Muthamizhan)
- tests: qa: split test\_tiering into smaller pieces ([pr#15146](#), Kefu Chai)

- tests: qa/suite: Added a smoke suite for ceph-ansible ([pr#12610](#), Tamil Muthamizhan)
- tests: qa/suite: replace reference to fs/xfs.yaml ([pr#14756](#), Yehuda Sadeh)
- tests: qa/suites/ceph-ansible: removing fs workunit ([pr#12928](#), Tamil Muthamizhan)
- tests: qa/suites/{ceph-ansible,rest}: OpenStack volumes ([pr#13672](#), Zack Cerza)
- tests: qa/suites/ceph-deploy: Drop OpenStack volume count ([pr#13706](#), Zack Cerza)
- tests: qa/suites: drop 'fs' facet, and add 'objectstore' facet where missing ([pr#14198](#), Sage Weil)
- tests: qa/suites: escape the parenthesis of the whitelist text ([pr#16722](#), Kefu Chai)
- tests: qa/suites: fix upgrade tests vs cluster full thrashing ([pr#13852](#), Sage Weil)
- tests: qa/suites/fs: Add openstack volume configuration ([pr#13640](#), Zack Cerza)
- tests: qa/suites/jewel-x/point-to-point: don't scan for keys on second s3tests either ([pr#14788](#), Sage Weil)
- tests: qa/suites/kcephfs: Openstack volume configuration ([pr#13634](#), Zack Cerza)
- tests: qa/suites/{knfs,hadoop,samba}: OpenStack volume configuration ([pr#13637](#), Zack Cerza)
- tests: qa/suites/krbd: Add openstack volume configuration ([pr#13631](#), Zack Cerza)
- tests: qa/suites/powercycle/osd/whitelist\_health: whitelist more ([pr#17306](#), Sage Weil)
- tests: qa/suites/powercycle: whitelist health for thrashing ([pr#16759](#), Sage Weil)
- tests: qa/suites/rados: a bit more whitelisting ([pr#16820](#), Sage Weil)
- tests: qa/suites/rados: fix ec thrashing ([pr#15087](#), Sage Weil)
- tests: qa/suites/rados/objectstore: enable experimental features for testing bluestore ([pr#13456](#), Kefu Chai, Dan Mick)
- tests: qa/suites/rados/singleton/all/erasure-code-nonregression: fix typo ([pr#16579](#), Sage Weil)
- tests: qa/suites/rados/singleton/all/mon-auth-caps: more osds so we can go clean ([pr#16225](#), Sage Weil)
- tests: qa/suites/rados/singleton-bluestore: concat settings ([pr#14884](#), Kefu Chai)

- tests: qa/suites/rados/singleton-nomsgr/all/multi-backfill-reject: sleep longer ([pr#16739](#), Sage Weil)
- tests: qa/suites/rados/singleton-nomsgr: fix syntax ([pr#15276](#), Sage Weil)
- tests: qa/suites/rados/thrash: make sure osds have map before legacy scrub ([pr#15117](#), Sage Weil)
- tests: qa/suites/rados/upgrade: restart mds ([pr#15517](#), Sage Weil)
- tests: qa/suites: Reduce fs combination tests for smoke, use bluestore ([pr#14854](#), Vasu Kulkarni)
- tests: qa/suites: Revert “qa/suites: add mon-reweight-min-pgs-per-osd = 4” ([pr#14584](#), Kefu Chai)
- tests: qa/suites/rgw: Add openstack volume configuration ([pr#13611](#), Zack Cerza)
- tests: qa/suites/upgarde/jewel-x/parallel: more whitelisting ([pr#16849](#), Sage Weil)
- tests: qa/suites/upgrade: add tiering test to hammer-jewel-x ([issue#19185](#), [pr#13805](#), Kefu Chai)
- tests: qa/suites/upgrade/hammer-jewel-x: add luminous.yaml ([issue#20342](#), [pr#15764](#), Kefu Chai)
- tests: qa/suites/upgrade/jewel-x: add mgr.x role ([pr#14689](#), Sage Weil)
- tests: qa/suites/upgrade/jewel-x: misc fixes for new health checks ([pr#16429](#), Sage Weil)
- tests: qa/suites/upgrade/kraken-x: do not thrash cluster full during upgrade ([issue#19232](#), [pr#13892](#), Dan Mick)
- tests: qa/suites/upgrade/kraken-x: misc fixes ([pr#14887](#), Sage Weil)
- tests: qa/suites/upgrade/kraken-x ([pr#13517](#), Sage Weil, Yuri Weinstein)
- tests: qa/suites/upgrade: set “sortbitwise” for jewel clusters ([pr#15661](#), Kefu Chai)
- tests: qa/suite/upgrade/jewel-x: various fixes ([pr#13734](#), Sage Weil)
- tests: qa/tasks: assert on pg status with a timeout ([issue#19594](#), [pr#14608](#), Kefu Chai)
- tests: qa/tasks/ceph: debug osd setup ([pr#16841](#), Sage Weil)
- tests: qa/tasks/ceph-deploy: create-keys explicitly ([pr#12867](#), Vasu Kulkarni)
- tests: qa/tasks/ceph-deploy: Fix bluestore options for ceph-deploy ([pr#16571](#),

Vasu Kulkarni)

- tests: qa/tasks/ceph-deploy: use the new create option during instantiation ([pr#12892](#), Vasu Kulkarni)
- tests: qa/tasks/ceph: don't hard-code cluster name when copying fsid ([pr#16212](#), Jason Dillaman)
- tests: qa/tasks/ceph\_manager: always fix pgp\_num when done with thrashosd task ([issue#19771](#), [pr#14931](#), Kefu Chai)
- tests: qa/tasks/ceph\_manager: 'ceph \$service tell ...' is obsolete ([pr#15252](#), Sage Weil)
- tests: qa/tasks/ceph.py: debug which pgs aren't scrubbing ([pr#13649](#), Sage Weil)
- tests: qa/tasks/ceph: raise exceptions if scrubbing fails or cannot proceed ([pr#15310](#), Sage Weil)
- tests: qa/tasks/ceph: should be "Waiting for all PGs", not "all osds" ([pr#16122](#), Kefu Chai)
- tests: qa/tasks/ceph: wait longer for scrub ([pr#16824](#), Sage Weil)
- tests: qa/tasks: few fixes to get ceph-deploy 1node to working state ([pr#14400](#), Vasu Kulkarni)
- tests: qa/tasks/radosbench: increase timeout ([pr#15885](#), Sage Weil)
- tests: qa/tasks/rebuild\_mondb: grant "mgr:allow \*" to client.admin ([issue#19439](#), [pr#14284](#), Kefu Chai)
- tests: qa/tasks/reg11184: use literal 'foo' instead pool\_name ([pr#16451](#), Kefu Chai)
- tests: qa/tasks/repair\_test: unset flags we set ([pr#15296](#), Sage Weil)
- tests: qa/tasks/rgw.py: start Apache before RadosGW ([pr#13846](#), Radoslaw Zarzynski)
- tests: qa/tasks/thrashosds-health.yaml: ignore MON\_DOWN ([issue#20910](#), [pr#17003](#), Sage Weil)
- tests: qa/tasks: use sudo to check ceph health for systemd test ([pr#14464](#), Vasu Kulkarni)
- tests: qa/tasks/workunit.py: use "overrides" as the default settings of workunit ([issue#19429](#), [pr#14281](#), Kefu Chai)
- tests: qa/tasks/workunit: use ceph.git as an alternative of ceph-ci.git for cloning workunit ([pr#13663](#), Kefu Chai)

- tests: qa/tasks/workunit: use the suite repo for cloning workunit ([pr#13452](#), Kefu Chai)
- tests: qa/tasks/workunit: use the suite repo for cloning workunit ([pr#13625](#), Kefu Chai)
- tests: qa/test\_rados\_tool.sh: POSIX dd only accepts 'k' as multiplier ([pr#12699](#), Willem Jan Withagen)
- tests: qa: timeout when waiting for mgr to be available in healthy() ([pr#16797](#), Josh Durgin)
- tests: qa: Using centos 7.2 for latest version ([pr#12806](#), Yuri Weinstein)
- tests: qa/workunits/ceph-helpers: display rejected string ([issue#20344](#), [pr#14468](#), Kefu Chai)
- tests: qa/workunits/ceph-helpers: enable experimental features for osd ([pr#16319](#), Kefu Chai)
- tests: qa/workunits/ceph-helpers.sh: use syntax understood by jq 1.3 ([pr#15530](#), Kefu Chai)
- tests: qa/workunits/ceph-helpers: test wait\_for\_health\_ok differently ([pr#16317](#), Kefu Chai)
- tests: qa/workunits/ceph-helpers: wait\_for\_clean() races with pg creation ([pr#12866](#), David Zafman)
- tests: qa/workunits/cephtool/test.sh: Be more liberal in testing health-output ([pr#14614](#), Willem Jan Withagen)
- tests: qa/workunits/cephtool/test.sh: “ceph osd stat” output changed, update accordingly ([pr#16444](#), Willem Jan Withagen, Kefu Chai)
- tests: qa/workunits/cephtool/test.sh: disable ‘fs status’ until bug is fixed ([issue#20761](#), [pr#16541](#), Sage Weil)
- tests: qa/workunits/cephtool/test.sh: fix test to watch audit channel ([pr#16470](#), Sage Weil)
- tests: qa/workunits/cephtool/test.sh: only include last line for epoch ([issue#20477](#), [pr#15770](#), Kefu Chai)
- tests: qa/workunits/rados/test.sh: print test name when it fails ([pr#13264](#), Kefu Chai)
- tests: rados: move cephtool.yaml to new singleton/bluestore subsuite ([issue#19797](#), [pr#14847](#), Nathan Cutler)
- tests: rbd/test\_lock\_fence.sh: fix rbdrw.py relative path ([issue#18388](#), [pr#12747](#),

- Nathan Cutler)
- tests: re-enable cephfs python tests on kclient ([issue#17193](#), [issue#18161](#), [pr#13200](#), Nathan Cutler)
  - tests: remove temporary file ([pr#12919](#), Kefu Chai)
  - tests: Revert “dummy: reduce run time, run user.yaml playbook” ([issue#18259](#), [pr#12506](#), Nathan Cutler)
  - tests: Revert “qa/tasks/workunit: use the suite repo for cloning workunit” ([pr#13495](#), Sage Weil)
  - tests: rgw.py: put client roles in a separate list ([issue#20417](#), [pr#15913](#), Nathan Cutler)
  - tests: rgw singleton: drop duplicate filestore-xfs.yaml ([pr#15959](#), Nathan Cutler)
  - tests: set -x in suites/iozone.sh workunit ([issue#19740](#), [pr#14713](#), Nathan Cutler)
  - tests: src/test/test\_denc.cc: Fix errors in buffer overflow ([pr#12653](#), Willem Jan Withagen)
  - tests: subst repo and branch in git.ceph.com URL in qa/tasks/cram.py and qa/tasks/qemu.py ([issue#18440](#), [pr#12816](#), Nathan Cutler)
  - tests: tasks/workunit.py: when cloning, use -depth=1 ([pr#14214](#), Dan Mick)
  - tests: test: add explicit braces to avoid ambiguous ‘else’ and to silence warnings ([pr#14472](#), Jos Collin)
  - tests: test: add override in test submodule ([pr#13773](#), liuchang0812)
  - tests: test: ceph osd stat out has changed, fix tests for that ([pr#16403](#), Willem Jan Withagen)
  - tests: test:Check make\_writeable() return value ([pr#15266](#), zhanglei)
  - tests: test: clean up unused variable ([pr#12873](#), liuchang0812)
  - tests: test/CMakeLists: disable test\_pidfile.sh ([issue#20975](#), [pr#17241](#), Sage Weil)
  - tests: test/compressor: disable isal tests if not available ([pr#14929](#), Kefu Chai)
  - tests: test: c\_read\_operations.cc: silence warning from -Wsign-compare ([pr#14888](#), Jos Collin)
  - tests: test: create asok files in a temp directory under \$TMPDIR ([issue#16895](#), [pr#16445](#), Kefu Chai)
  - tests: test/crush: silence warnings from -Walloc-size-larger-than= and -

- Wstringop-overflow ([pr#15173](#), Jos Collin)
- tests: test: c\_write\_operations.cc: silence warning from -Wsign-compare ([pr#14889](#), Jos Collin)
  - tests: test: Division by zero in Legacy::encode\_n() ([pr#15902](#), Jos Collin)
  - tests: test/encoding: fix readable.sh bugs; fix ceph-object-corpus ([pr#13678](#), Sage Weil)
  - tests: test/fio\_ceph\_objectstore: fix fio plugin build failure by engine\_data ([pr#15044](#), lisali)
  - tests: test/fio: Fix assert in set\_cache\_shards in bluestore fio ([pr#15659](#), Xiaoyan Li)
  - tests: test/fio: fix lack of setting for Sequencer::shard\_hint ([pr#15571](#), Igor Fedotov)
  - tests: test/fio: print all perfcounters rather than objectstore itself ([pr#16339](#), Jianpeng Ma)
  - tests: test/fio: remove experimental option for bluestore & rocksdb ([pr#16263](#), Pan Liu)
  - tests: test: Fixes for test\_pidfile ([issue#20770](#), [pr#16587](#), David Zafman)
  - tests: test: fixing assert that creates warning: comparison between signed and unsigned integer expressions ([pr#14794](#), Jos Collin)
  - tests: test: Fix mismatched sign comparison in histogram test ([pr#13362](#), Adam C. Emerson)
  - tests: test: Fix reg11184 test to remove extraneous pg ([pr#16265](#), David Zafman)
  - tests: test: fix test\_pidfile ([pr#13646](#), yaoning)
  - tests: test/fsx: Remove the dead code associated with aio backend ([pr#14905](#), Zhou Zhengping)
  - tests: test: Initialize pointer variables in TestMemIoCtxImpl ([pr#16785](#), amitkuma)
  - tests: test/librados: Initialize member variables in aio.cc ([pr#16845](#), amitkuma)
  - tests: test: librados\_test\_stub: tmap\_update: return -ENOENT when removing nonexistent key ([pr#12667](#), Mykola Golub)
  - tests: test: migrate atomic\_t to std::atomic ([pr#14655](#), Jesse Williamson)
  - tests: test,mon,msg: kill clang analyzer warnings ([pr#16320](#), Kefu Chai)

- tests: test/mon: silence warnings from -Wreorder ([pr#15692](#), Jos Collin)
- tests: test/msgr: fixed the hang issue for perf\_msg\_client ([pr#16358](#), Pan Liu)
- tests: test/msgr: silence warnings from -Wsign-compare ([pr#15356](#), Jos Collin)
- tests: test/msgr: silence warnings from -Wsign-compare ([pr#15570](#), Jos Collin)
- tests: test/objectstore: chain\_xattr: fix wrong memset usage to fill buf ([pr#14277](#), Weibing Zhang)
- tests: test/objectstore: Check apply\_transaction() return values ([pr#15171](#), zanglei)
- tests: test/objectstore/: Check put\_ref return value ([pr#15007](#), zanglei)
- tests: test/old: Removed commented code ([pr#15366](#), Jos Collin)
- tests: test/osdc: fix comparison error and silence warning from -Wunused-value ([pr#15353](#), Willem Jan Withagen)
- tests: test/osd: kill compile warning ([pr#16669](#), Yan Jun)
- tests: test/osd/osd-dup.sh: lower wb fd throttle limits ([pr#14984](#), Dan Mick)
- tests: test/osd/osd-dup.sh: use wait\_for\_clean ([pr#15722](#), Dan Mick)
- tests: test/osd/osd-scrub-repair.sh: disable ec\_overwrite tests on FreeBSD ([pr#15445](#), Willem Jan Withagen)
- tests: test/osd/osd-scrub-repair.sh: Fix diff options on FreeBSD ([pr#15914](#), Willem Jan Withagen)
- tests: test/osd: Removed Commented Code - 2 ([issue#20207](#), [pr#15540](#), Jos Collin)
- tests: test: osd/TestOSDMap.cc: fix Clang complain about promotion ([pr#15525](#), Willem Jan Withagen)
- tests: test/rados: fix wrong parameter order of RETURN1\_IF\_NOT\_VAL ([pr#16589](#), Yan Jun)
- tests: test: reg11184 might not always find pg 2.0 prior to import ([pr#16610](#), David Zafman)
- tests: test: Rename FileJournal object to distinguish ([pr#15279](#), Jos Collin)
- tests: test: replace hard-code binary names with variables ([pr#12675](#), liuchang0812)
- tests: test: sed on FreeBSD requires “-i extension”, so use gsed ([pr#13903](#), Willem Jan Withagen)

- tests: test: s/osd\_objectstore\_type/osd\_objectstore ([pr#16469](#), xie xingguo)
- tests: test: test\_denc.cc: silence warning from -Wsign-compare ([pr#15355](#), Jos Collin)
- tests: test: Test fix for SnapSet change ([pr#15161](#), David Zafman)
- tests: test: test\_pidfile running 2nd mon has unreliable log output ([pr#16635](#), David Zafman)
- tests: test: Thrasher: do not update pools\_to\_fix\_pgp\_num if nothing happens ([pr#13518](#), Kefu Chai)
- tests: test: Thrasher: update pgp\_num of all expanded pools if not yet ([pr#13367](#), Kefu Chai)
- tests: test/unittest\_bluefs: check whether mounted success ([pr#14988](#), shiqi)
- tests: test/unittest\_bluefs: remove unused variable ([pr#14006](#), shiqi)
- tests: test: unittest\_hostname compile error on freebsd ([pr#13739](#), liuchang0812)
- tests: test: update test\_rados\_tool.sh, use POOL and OBJ var ([pr#12706](#), liuchang0812)
- tests: test: use 7130 for crush-classes.sh ([pr#14783](#), Loic Dachary)
- tests: test/vstart\_wrapper.sh: display\_log on test failure ([pr#15620](#), Kefu Chai)
- tests: test: warning: comparison between signed and unsigned integer expressions ([pr#14705](#), Jos Collin)
- tests: Thrasher: eliminate a race between kill\_osd and \_\_init\_\_ ([issue#18799](#), [pr#13237](#), Nathan Cutler)
- tests: Thrasher: handle “OSD has the store locked” gracefully ([issue#19556](#), [pr#14415](#), Nathan Cutler)
- tests,tools: script/find\_dups\_in\_pg\_log: scrip to find dup requests due to short pg logs ([pr#13417](#), Sage Weil)
- tests,tools: test, ceph-osdomap-tool: kill clang warnings ([pr#15905](#), Kefu Chai)
- tests,tools: test: kill warnings ([pr#14892](#), Kefu Chai)
- tests: update SUSE yaml facets in qa/distros/all ([issue#18856](#), [pr#13313](#), Nathan Cutler)
- tests: workunit: request branch when cloning ([pr#14260](#), Kefu Chai, Dan Mick)
- tools: add override in tool submodule ([pr#13776](#), liuchang0812)

- tools: brag: count the number of mds in fsmap not in mdsmap ([issue#19192](#), [pr#13798](#), Peng Zhang)
- tools: ceph\_common.sh: fix syntax error ([issue#17826](#), [pr#13419](#), Dan Mick)
- tools: ceph-conf: fix typo in usage: 'mon add' should be 'mon addr' ([pr#15935](#), Peng Zhang)
- tools: ceph-create-keys: add an argument to override default 10-minute timeout ([pr#16049](#), Douglas Fuller)
- tools: ceph-detect-init: adding Arch Linux support ([pr#12787](#), Jamin W. Collins)
- tools: ceph-detect-init: Adds Oracle Linux Server and Oracle VM Server detect ([pr#13917](#), Nikita Gerasimov)
- tools: ceph-detect-init: detect init system by poking the system ([issue#19884](#), [pr#15043](#), Kefu Chai)
- tools: ceph-disk: Add fix subcommand ([pr#13310](#), Boris Ranto)
- tools: ceph-disk: change the lockbox partition number to 5 ([issue#20556](#), [pr#16247](#), Shangzhong Zhu)
- tools: ceph-disk: command invocation needs all fields separate ([pr#15733](#), Willem Jan Withagen)
- tools: ceph-disk: convert none str to str before printing it ([issue#18371](#), [pr#12760](#), Kefu Chai)
- tools: ceph-disk: do not remove mount point if deactivate -once ([pr#16474](#), Song Shun)
- tools: ceph-disk: Fix for missing 'not' in \*\_is\_diskdevice checks ([issue#20706](#), [pr#16481](#), Nikita Gerasimov)
- tools: ceph\_disk/main.py: FreeBSD root has wheel for group ([pr#16609](#), Willem Jan Withagen)
- tools: ceph-disk: s/ceph\_osd\_mkfs/command\_check\_call/ ([issue#20685](#), [pr#16427](#), Zhu Shangzhong)
- tools: ceph.in: add help for locally-handled commands ([pr#13288](#), Dan Mick)
- tools: ceph.in: adjust usage width according to user's tty ([pr#15190](#), Kefu Chai)
- tools: ceph.in: assert(state==connected) before help\_for\_target() ([pr#15156](#), Kefu Chai)
- tools: ceph.in: drop the compatibility to handle non json commands ([pr#15508](#), Kefu Chai)

- tools: ceph.in: filter out audit from ceph -w ([pr#16345](#), John Spray)
- tools: ceph.in, mgr: misc cleanups ([pr#16229](#), liuchang0812)
- tools: ceph.in: print return code when json\_command failed ([pr#15378](#), liuchang0812)
- tools: ceph-objectstore-tool: Handle object names that are also valid json ([pr#12848](#), David Zafman)
- tools: ceph-release-notes: escape asterisks not for inline emphasis ([pr#16199](#), Kefu Chai)
- tools: ceph-release-notes: escape \_ for unintended links ([issue#17499](#), [pr#16528](#), Kefu Chai)
- tools: ceph-release-notes: handle an edge case ([pr#16277](#), Nathan Cutler)
- tools: ceph-release-notes: ignore low-numbered PRs ([issue#18695](#), [pr#13151](#), Nathan Cutler)
- tools: ceph-release-notes: port it to py3 ([pr#16261](#), Kefu Chai)
- tools: ceph-release-notes: prefixes and pep8 compliance ([pr#14156](#), Nathan Cutler)
- tools: ceph-release-notes: refactor and fix regressions ([pr#16411](#), Nathan Cutler)
- tools: ceph-release-notes: strip trailing punctuation ([pr#14385](#), Nathan Cutler)
- tools: ceph-rest-api: be more tolerant on network failure ([issue#20115](#), [pr#15706](#), Kefu Chai)
- tools: ceph-volume: adds functional CI testing #16919 ([pr#16970](#), Andrew Schoen, Alfredo Deza)
- tools: ceph-volume docs ([pr#17124](#), Alfredo Deza)
- tools: ceph-volume: initial take on ceph-volume CLI tool ([pr#16632](#), Dan Mick, Alfredo Deza)
- tools: ceph-volume: Use a delimited CLI output parser instead of JSON ([pr#17123](#), Alfredo Deza)
- tools: change compare\_exchange\_weak to compare\_exchange\_strong ([pr#15030](#), Jesse Williamson)
- tools: Cleanup dead code in ceph-objectstore-tool ([pr#15812](#), David Zafman)
- tools: fio\_ceph\_objectstore: Print db\_statistics when rocksdb\_perf is enabled ([pr#15796](#), Xiaoyan Li)
- tools: init-ceph: print trailing n in "status" output ([pr#13351](#), Kefu Chai)

- tools: init-ceph: should have a space before "]" ([pr#14796](#), Kefu Chai)
- tools: os/bluestore/bluestore\_tool: add sanity check to get rid of occasionally crash ([pr#16013](#), xie xingguo)
- tools: rados: check for negative return value of rados\_create\_with\_context() as its comment put ([pr#10893](#), zhang.zezhu)
- tools: rados: fix typo in 'df' column name ([pr#15603](#), Ilya Dryomov)
- tools: rados: out json 'df' values as numbers, not strings ([issue#15546](#), [pr#14644](#), Sage Weil)
- tools: script: add docker core dump debugger ([pr#16375](#), Patrick Donnelly)
- tools: script: ceph-release-notes check orig. issue only for backports ([pr#12979](#), Abhishek Lekshmanan)
- tools: script/sephia\_bt.sh: no need to pass version and sha1 anymore ([pr#13380](#), Kefu Chai)
- tools: src/ceph-disk/ceph\_disk/main.py: Make 'ceph-disk list' work on FreeBSD ([pr#14483](#), Willem Jan Withagen)
- tools: stop.sh: boilerplate error (don't stop mon when stopping mgr) ([pr#14461](#), Dan Mick)
- tools: support hammer in rbd\_recover\_tool ([pr#12413](#), Bartłomiej Święcki)
- tools: tools/ceph\_kvstore\_tool: add "bluestore-kv" to usage ([pr#15326](#), xie xingguo)
- tools: tools/crushtool: replicated-rule API support ([pr#15011](#), xie xingguo)
- tools: tools/rados: add a parameter "-offset" to rados put command ([pr#12674](#), liuchang0812)
- tools: tools/rados: Check return value of connect ([issue#19319](#), [pr#14057](#), Brad Hubbard)
- tools: tools/rados: fixed typo in help information ([pr#15618](#), Pan Liu)
- tools: tools/rados: remove useless function declaration ([pr#12566](#), liuchang0812)
- tools: tools/rados: some cleanups ([pr#16147](#), Yan Jun)
- tools: vstart: "debug\_ms=1" for mgr by default ([pr#15127](#), Kefu Chai)
- tools: vstart: don't configure rgw\_dns\_name ([pr#13411](#), Yehuda Sadeh)
- tools: vstart: don't create cluster by default ([pr#13891](#), Yehuda Sadeh)

- tools: vstart: print "start osd.\$id" instead of "start osd\$id" ([pr#15427](#), Kefu Chai)
- tools: vstart.sh: bind restful, dashboard to ::, not 127.0.0.1 ([pr#16349](#), Sage Weil)
- tools: vstart.sh: do not add host for mgr.\* section if not \$overwrite\_conf ([pr#13767](#), Kefu Chai)
- tools: warning, '%.16x' directive output truncated writing 16 bytes into a region of size 9 ([pr#14292](#), Jos Collin)
- tracing: don't include oid when tracing at dequeue\_op() ([pr#13410](#), Yehuda Sadeh)

## v11.2.1 Kraken

This is the first bugfix release for Kraken, and probably the last release of the Kraken series (Kraken will be declared “End Of Life” (EOL) when Luminous is declared stable). It contains a large number of bugfixes across all Ceph components.

We recommend that all v11.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

- In previous versions, if a client sent an op to the wrong OSD, the OSD would reply with ENXIO. The rationale here is that the client or OSD is clearly buggy and we want to surface the error as clearly as possible. We now only send the ENXIO reply if the `osd_enxio_on_misdirected_op` option is enabled (it’s off by default). This means that a VM using librbd that previously would have gotten an EIO and gone read-only will now see a blocked/hung IO instead.
- There was a bug introduced in Jewel (#19119) that broke the mapping behavior when an “out” OSD that still existed in the CRUSH map was removed with ‘`osd rm`’. This could result in ‘misdirected op’ and other errors. The bug is now fixed, but the fix itself introduces the same risk because the behavior may vary between clients and OSDs. To avoid problems, please ensure that all OSDs are removed from the CRUSH map before deleting them. That is, be sure to do:

```
1. ceph osd crush rm osd.123
```

before:

```
1. ceph osd rm osd.123
```

- This release greatly improves control and throttling of the snap trimmer. It introduces the “`osd max trimming pgs`” option (defaulting to 2), which limits how many PGs on an OSD can be trimming snapshots at a time. And it restores the safe use of the “`osd snap trim sleep`” option, which defaults to 0 but otherwise adds the given number of seconds in delay between every dispatch of trim operations to the underlying system.

## Other Notable Changes

- build/ops: ceph-base missing dependency for psmisc in Ubuntu Xenial ([issue#19129](#), [issue#19564](#), [pr#14425](#), Nathan Cutler)

- build/ops: logrotate is missing from debian package (kraken, master) ([issue#19670](#), [issue#19390](#), [pr#14734](#), Kefu Chai)
- build/ops: selinux: Do parallel relabel on package install ([issue#20077](#), [issue#20184](#), [issue#20191](#), [issue#20193](#), [pr#15509](#), Boris Ranto)
- build/ops: spec file mentions non-existent ceph-create-keys systemd unit file, causing ceph-mon units to not be enabled via preset ([issue#19460](#), [pr#14315](#), Sébastien Han)
- build/ops: systemd restarts Ceph Mon to quickly after failing to start ([issue#18635](#), [issue#18721](#), [pr#13185](#), Wido den Hollander)
- build/ops: systemd: Start OSDs after MONs ([issue#18907](#), [issue#18516](#), [pr#13494](#), Boris Ranto)
- ceph-disk: Add fix subcommand kraken back-port ([issue#19544](#), [pr#14345](#), Boris Ranto)
- ceph-disk: does not support cluster names different than 'ceph' ([issue#18973](#), [issue#17821](#), [pr#13497](#), Loic Dachary)
- ceph-disk: enable directory backed OSD at boot time ([issue#19628](#), [issue#19647](#), [pr#14604](#), Loic Dachary)
- ceph-disk: error on \_bytes2str ([issue#18431](#), [issue#18371](#), [pr#13501](#), Kefu Chai)
- ceph-disk: fails if OSD udev rule triggers prior to mount of /var ([issue#20150](#), [issue#19941](#), [pr#16092](#), Loic Dachary)
- ceph-disk: Fix getting wrong group name when -setgroup in bluestore ([issue#18956](#), [pr#13488](#), craigchi)
- ceph-disk list reports mount error for OSD having mount options with SELinux context ([issue#19537](#), [issue#17331](#), [pr#14403](#), Brad Hubbard)
- ceph-disk prepare get wrong group name in bluestore ([issue#18997](#), [pr#13543](#), craigchi)
- ceph-disk: Racing between partition creation & device node creation ([issue#20034](#), [pr#16138](#), Erwan Velu)
- ceph-disk: separate ceph-osd -check-needs-\* logs ([issue#20010](#), [issue#19888](#), [pr#16135](#), Loic Dachary)
- cephfs: buffer overflow in test LibCephFS.DirLs ([issue#18941](#), [issue#19045](#), [pr#14571](#), "Yan, Zheng")
- cephfs: ceph-fuse crash during snapshot tests ([issue#18552](#), [issue#18460](#), [pr#14563](#), Yan, Zheng)

- cephfs: ceph-fuse does not recover after lost connection to MDS ([issue#19678](#), [issue#18757](#), [pr#16105](#), Henrik Korkuc)
- cephfs: client: fix the cross-quota rename boundary check conditions ([issue#18700](#), [pr#14567](#), Greg Farnum)
- cephfs: Deadlock on two ceph-fuse clients accessing the same file ([issue#20028](#), [issue#19635](#), [pr#16191](#), "Yan, Zheng")
- cephfs: fragment space check can cause replayed request fail ([issue#18660](#), [issue#18706](#), [pr#14568](#), "Yan, Zheng")
- cephfs: MDS crashes on missing metadata object ([issue#18179](#), [issue#18566](#), [pr#14565](#), Yan, Zheng)
- cephfs: MDS heartbeat timeout during rejoin, when working with large amount of caps/inodes ([issue#19118](#), [issue#19335](#), [pr#14572](#), John Spray)
- cephfs: mds is crushed, after I set about 400 64KB xattr kv pairs to a file ([issue#19674](#), [issue#19033](#), [pr#16103](#), Yang Honggang)
- cephfs: MDS server crashes due to inconsistent metadata ([issue#19406](#), [issue#19620](#), [pr#14574](#), John Spray)
- cephfs: mds/StrayManager: avoid reusing deleted inode in StrayManager::\_purge\_stray\_logged ([issue#18950](#), [pr#14570](#), Zhi Zhang)
- cephfs: mount point break off problem after mds switch ([issue#19667](#), [issue#19437](#), [pr#16100](#), Guan yunfei, Sage Weil)
- cephfs: non-local quota changes not visible until some IO is done ([issue#17939](#), [issue#19763](#), [pr#16108](#), John Spray)
- cephfs: No output for ceph mds rmfailed 0 -yes-i-really-mean-it command ([issue#19483](#), [issue#16709](#), [pr#14573](#), John Spray)
- cephfs: normalize file open flags internally used by cephfs ([issue#19845](#), [pr#14998](#), Jan Fajerski)
- cephfs: segfault in handle\_client\_caps ([issue#18306](#), [issue#18616](#), [pr#14566](#), Yan, Zheng)
- cephfs: speed up readdir by skipping unwanted dn ([issue#18531](#), [pr#13028](#), Xiaoxi Chen)
- cephfs: src/test/pybind/test\_cephfs.py fails ([issue#20500](#), [issue#19890](#), [pr#16114](#), "Yan, Zheng")
- cephfs: test\_client\_recovery.TestClientRecovery fails ([issue#18562](#), [issue#18396](#), [pr#14564](#), Yan, Zheng)

- cephfs test failures (ceph.com/qa is broken, should be download.ceph.com/qa) ([issue#18574](#), [issue#18604](#), [pr#13024](#), John Spray)
- cephfs: Test failure: test\_data\_isolated  
(`tasks.cephfs.test_volume_client.TestVolumeClient`) ([issue#18914](#), [issue#19676](#), [pr#16104](#), "Yan, Zheng")
- cephfs: test\_open\_inode fails ([issue#18899](#), [issue#18661](#), [pr#14569](#), John Spray)
- client: populate metadata during mount ([issue#18361](#), [issue#18540](#), [pr#12951](#), John Spray)
- client: segfault on ceph\_rmdir path / ([issue#18612](#), [issue#9935](#), [pr#13030](#), Michal Jarzabek)
- cls\_rbd: default initialize snapshot namespace for legacy clients ([issue#19413](#), [issue#19833](#), [pr#14934](#), Jason Dillaman)
- cls/rgw: list\_plain\_entries() stops before bi\_log entries ([issue#19876](#), [issue#20015](#), [pr#15384](#), Casey Bodley)
- common: monitor creation with IPv6 public network segfaults ([issue#19465](#), [issue#19371](#), [pr#14323](#), Fabian Grünbichler)
- common: possible lockdep false alarm for ThreadPool lock ([issue#18819](#), [issue#18894](#), [pr#13487](#), Mykola Golub)
- core: api\_misc: [ FAILED ] LibRadosMiscConnectFailure.ConnectFailure ([issue#19561](#), [issue#15368](#), [pr#14733](#), Sage Weil)
- core: bluestore bdev: flush no-op optimization is racy ([issue#20495](#), [issue#19326](#), [issue#19327](#), [issue#19250](#), [issue#19251](#), [pr#14736](#), Sage Weil)
- core: improve control and throttling of the snap trimmer ([issue#19329](#), [issue#19931](#), [pr#14597](#), Samuel Just, Greg Farnum)
- core: two instances of omap\_digest mismatch ([issue#19391](#), [pr#14200](#), Samuel Just, David Zafman)
- doc: PendingReleaseNotes: warning about 'osd rm ...' and #13733 ([issue#19119](#), [pr#14506](#), Sage Weil)
- doc: Python Swift client commands in Quick Developer Guide don't match configuration in vstart.sh ([issue#17746](#), [issue#18571](#), [pr#13044](#), Ronak Jain)
- doc: rgw: admin ops: fix the quota section ([issue#19397](#), [issue#19462](#), [pr#14521](#), Chu, Hua-Rong)
- fix: rgw crashed caused by shard id out of range when listing data log ([issue#20156](#), [issue#19732](#), [pr#16173](#), redickwang)

- fuse: TestVolumeClient.test\_evict\_client failure creating pidfile ([issue#18439](#), [issue#18309](#), [pr#12813](#), Nathan Cutler)
- librbd: allow to open an image without opening parent image ([issue#18609](#), [issue#18325](#), [pr#13132](#), Ricardo Dias)
- librbd: corrected resize RPC message backwards compatibility ([issue#19636](#), [issue#19659](#), [pr#14620](#), Jason Dillaman)
- librbd: Incomplete declaration for ContextWQ in librbd/Journal.h ([issue#18862](#), [issue#18892](#), [pr#14153](#), Boris Ranto)
- librbd: is\_exclusive\_lock\_owner API should ping OSD ([issue#19467](#), [issue#19287](#), [pr#14480](#), Jason Dillaman)
- librbd: possible race in ExclusiveLock handle\_peer\_notification ([issue#19368](#), [pr#14163](#), Mykola Golub)
- librbd: prevent self-blacklisting during break lock ([issue#18703](#), [issue#18666](#), [pr#13201](#), Jason Dillaman)
- make check fails with Error EIO: load dlopen(build/lib/libec\_FAKE.so): build/lib/libec\_FAKE.so: cannot open shared object file: No such file or directory ([issue#20487](#), [issue#20345](#), [issue#18876](#), [pr#16069](#), Kefu Chai, Kyr Shatskyy)
- mds: assert fail when shutting down ([issue#19672](#), [issue#19204](#), [pr#16102](#), John Spray)
- mds: C\_MDSInternalNoop::complete doesn't free itself ([issue#19664](#), [issue#19501](#), [pr#16099](#), "Yan, Zheng")
- mds: daemon goes readonly writing backtrace for a file whose data pool has been removed ([issue#19669](#), [issue#19401](#), [pr#16101](#), John Spray)
- mds: damage reporting by ino number is useless ([issue#18509](#), [issue#19680](#), [pr#16106](#), John Spray)
- mds: Decode errors on backtrace will crash MDS ([issue#18311](#), [issue#18463](#), [pr#12835](#), John Spray)
- mds: enable daemon to start when session ino info is corrupt ([issue#19710](#), [issue#16842](#), [pr#16107](#), John Spray)
- mds: failed filelock.can\_read(-1) assertion in Server::\_dir\_is\_nonempty ([issue#18707](#), [issue#18578](#), [pr#13555](#), Yan, Zheng)
- mds: finish clientreplay requests before requesting active state ([issue#18678](#), [issue#18461](#), [pr#13112](#), Yan, Zheng)
- mds: unresponsive when truncating a very large file ([issue#19755](#), [issue#20026](#),

- pr#16190, "Yan, Zheng")
- mon: cache tiering: base pool last\_force\_resend not respected (racing read got wrong version) ([issue#18366](#), [issue#18403](#), [pr#13116](#), Sage Weil)
- mon crash on shutdown, lease\_ack\_timeout event ([issue#19928](#), [issue#19825](#), [pr#15084](#), Kefu Chai, Alexey Sheplyakov)
- mon: fail to form large quorum; msg/async busy loop ([issue#20230](#), [issue#20315](#), [pr#15729](#), Haomai Wang)
- mon: force\_create\_pg could leave pg stuck in creating state ([issue#19181](#), [issue#18298](#), [pr#13790](#), Adam C. Emerson, Sage Weil)
- mon/MonClient: make get\_mon\_log\_message() atomic ([issue#19618](#), [issue#19427](#), [pr#14588](#), Kefu Chai)
- mon: 'osd crush move ...' doesnt work on osds ([issue#18682](#), [issue#18587](#), [pr#13500](#), Sage Weil)
- mon: osd crush set crushmap need sanity check ([issue#19302](#), [issue#20365](#), [pr#16143](#), Loic Dachary)
- mon: peon wrongly delete routed pg stats op before receive pg stats ack ([issue#18554](#), [issue#18458](#), [pr#13046](#), Mingxin Liu)
- mon/PGMap: factor mon\_osd\_full\_ratio into MAX AVAIL calc ([issue#18522](#), [issue#20035](#), [pr#15237](#), Sage Weil)
- msg/simple/SimpleMessenger.cc: 239: FAILED assert(!cleared) ([issue#15784](#), [issue#18378](#), [pr#16133](#), Sage Weil)
- multisite: rest api fails to decode large period on 'period commit' ([issue#19505](#), [issue#19616](#), [issue#19614](#), [issue#20244](#), [issue#19488](#), [issue#19776](#), [issue#20293](#), [issue#19746](#), [pr#16161](#), Casey Bodley, Abhishek Lekshmanan)
- objecter: full\_try behavior not consistent with osd ([issue#19560](#), [issue#19430](#), [pr#14732](#), Sage Weil)
- objecter: epoch\_barrier isn't respected in \_op\_submit() ([issue#19396](#), [issue#19496](#), [pr#14331](#), Ilya Dryomov)
- os/bluestore: deep decode onode value ([issue#20366](#), [pr#15792](#), Sage Weil)
- os/bluestore: fix Allocator::allocate() int truncation ([issue#20884](#), [issue#18595](#), [pr#13011](#), Sage Weil)
- osd: allow client throttler to be adjusted on-fly, without restart ([issue#18791](#), [issue#18793](#), [pr#13216](#), Piotr Dałek)
- osd: An OSD was seen getting ENOSPC even with osdfailsafe\_full\_ratio passed

- ([issue#20544](#), [issue#16878](#), [issue#19340](#), [issue#19841](#), [issue#20672](#), [pr#16134](#), Sage Weil, David Zafman)
- osd: bogus assert when checking acting set on recovery completion in rados/upgrade ([issue#18999](#), [pr#13542](#), Sage Weil)
  - osd: calc\_clone\_subsets misuses try\_read\_lock vs missing ([issue#18610](#), [issue#18583](#), [issue#18723](#), [issue#17831](#), [pr#14616](#), Samuel Just)
  - osd: ceph degraded and misplaced status output inaccurate ([issue#18619](#), [issue#19480](#), [pr#14322](#), David Zafman)
  - osd: condition object\_info\_t encoding on required (not up) features ([issue#18842](#), [issue#18831](#), [issue#18814](#), [pr#13485](#), Ilya Dryomov)
  - osd: do not send ENXIO on misdirected op by default ([issue#19622](#), [pr#13253](#), Sage Weil)
  - osd: FAILED assert(object\_contexts.empty()) (live on master only from Jan-Feb 2017, all other instances are different) ([issue#20522](#), [issue#20523](#), [issue#18927](#), [issue#18809](#), [pr#16132](#), Samuel Just)
  - osd: -flush-journal: sporadic segfaults on exit ([issue#18952](#), [issue#18820](#), [pr#13490](#), Alexey Sheplyakov)
  - osd: Give requested scrubs a higher priority ([issue#19685](#), [issue#15789](#), [pr#14735](#), David Zafman)
  - osd: Implement asynchronous scrub sleep ([issue#20033](#), [issue#19986](#), [issue#20173](#), [issue#19497](#), [pr#15526](#), Brad Hubbard)
  - osd: leaked MOSDMap ([issue#19760](#), [issue#18293](#), [pr#14942](#), Sage Weil)
  - osd: leveldb corruption leads to Operation not permitted not handled and assert ([issue#18037](#), [issue#18418](#), [pr#12790](#), Nathan Cutler)
  - osd: metadata reports filestore when using bluestore ([issue#18677](#), [issue#18638](#), [pr#16083](#), Wido den Hollander)
  - osd: New added OSD always down when full flag is set ([issue#19485](#), [pr#14321](#), Mingxin Liu)
  - osd: Object level shard errors are tracked and used if no auth available ([issue#20089](#), [pr#15421](#), David Zafman)
  - osd: os/bluestore: fix statfs to not include DB partition in free space ([issue#18599](#), [issue#18722](#), [pr#13284](#), Sage Weil)
  - osd: osd/PrimaryLogPG: do not call on\_shutdown() if (pg.deleting) ([issue#19902](#), [issue#19916](#), [pr#15066](#), Kefu Chai)

- osd: pg log split does not rebuild index for parent or child ([issue#19315](#), [issue#18975](#), [pr#14048](#), Sage Weil)
- osd: pglog: with config, don't assert in the presence of stale diverg... ([issue#17916](#), [issue#19702](#), [pr#14646](#), Greg Farnum)
- osd: publish PG stats when backfill-related states change ([issue#18497](#), [issue#18369](#), [pr#13295](#), Sage Weil)
- osd: Revert "PrimaryLogPG::failed\_push: update missing as well" ([issue#18659](#), [pr#13091](#), David Zafman)
- osd: unlock sdata\_op\_ordering\_lock with sdata\_lock hold to avoid missing wakeup signal ([issue#20443](#), [pr#15962](#), Alexey Sheplyakov)
- pre-jewel "osd rm" incrementals are misinterpreted ([issue#19209](#), [issue#19119](#), [pr#13883](#), Ilya Dryomov)
- rbd: Add missing parameter feedback to 'rbd snap limit' ([issue#18601](#), [pr#14537](#), Tang Jin)
- rbd: [api] is\_exclusive\_lock\_owner shouldn't return -EBUSY ([issue#20266](#), [issue#20182](#), [pr#16187](#), Jason Dillaman)
- rbd: [api] temporarily restrict (rbd\_)mirror\_peer\_add from adding multiple peers ([issue#19256](#), [issue#19324](#), [pr#14545](#), Jason Dillaman)
- rbd: attempting to remove an image with incompatible features results in partial removal ([issue#18456](#), [issue#18315](#), [pr#13247](#), Dongsheng Yang)
- rbd: [cli] ensure positional arguments exist before casting ([issue#20264](#), [issue#20185](#), [pr#16186](#), Jason Dillaman)
- rbd: cli: map with cephx disabled results in error message ([issue#19035](#), [issue#20517](#), [pr#16298](#), Jason Dillaman)
- rbd: [ FAILED ] TestJournalTrimmer.RemoveObjectsWithOtherClient ([issue#18769](#), [issue#18738](#), [pr#14147](#), Jason Dillaman)
- rbd: Improve compatibility between librbd + krbd for the data pool ([issue#18771](#), [issue#18653](#), [pr#14539](#), Jason Dillaman)
- rbd: Issues with C API image metadata retrieval functions ([issue#19588](#), [issue#19611](#), [pr#15612](#), Mykola Golub)
- rbd: 'metadata\_set' API operation should not change global config setting ([issue#18465](#), [issue#18549](#), [pr#14534](#), Mykola Golub)
- rbd-mirror: additional test stability improvements ([issue#18935](#), [issue#18947](#), [pr#14155](#), Jason Dillaman)

- rbd-mirror: deleting a snapshot during sync can result in read errors ([issue#19037](#), [issue#18990](#), [pr#14622](#), Jason Dillaman)
- rbd-mirror: ensure missing images are re-synced when detected ([issue#20022](#), [issue#19811](#), [pr#15486](#), Jason Dillaman)
- rbd-mirror: failover and fallback of unmodified image results in split-brain ([issue#19872](#), [issue#19858](#), [pr#14974](#), Jason Dillaman)
- rbd-mirror: potential race mirroring cloned image ([issue#18501](#), [issue#17993](#), [pr#14533](#), Jason Dillaman)
- rbd-mirror: sporadic image replayer shut down failure ([issue#18493](#), [issue#18441](#), [pr#14531](#), Jason Dillaman)
- rbd-nbd: add signal handler ([issue#19621](#), [issue#19349](#), [pr#16098](#), Kefu Chai, Pan Liu)
- rbd-nbd: check /sys/block/nbdX/size to ensure kernel mapped correctly ([issue#18970](#), [issue#17951](#), [issue#18910](#), [issue#18335](#), [pr#14540](#), Mykola Golub, Pan Liu)
- rbd: Possible deadlock performing a synchronous API action while refresh in-progress ([issue#18495](#), [issue#18419](#), [pr#14532](#), Jason Dillaman)
- rbd: Potential IO hang if image is flattened while read request is in-flight ([issue#19832](#), [issue#20154](#), [pr#16184](#), Jason Dillaman)
- rbd: [qa] crash in journal-enabled fsx run ([issue#18618](#), [issue#18632](#), [pr#14538](#), Jason Dillaman)
- rbd: qemu crash triggered by network issues ([issue#18776](#), [issue#18436](#), [pr#13245](#), Jason Dillaman)
- rbd: 'rbd bench-write' will crash if -io-size is 4G ([issue#18422](#), [issue#18557](#), [pr#14536](#), Gaurav Kumar Garg)
- rbd: rbd\_clone\_copy\_on\_read ineffective with exclusive-lock ([issue#19173](#), [issue#18888](#), [pr#14543](#), Venky Shankar)
- rbd: rbd -pool=x rename y z does not work ([issue#18777](#), [issue#18326](#), [pr#14149](#), Gaurav Kumar Garg)
- rbd: refuse to use an ec pool that doesn't support overwrites ([issue#19081](#), [issue#19336](#), [pr#16096](#), Jason Dillaman)
- rgw: add apis to support ragweed suite ([issue#19809](#), [pr#14852](#), Yehuda Sadeh)
- rgw: add the remove-x-delete feature to cancel swift object expiration ([issue#19472](#), [issue#19074](#), [pr#14522](#), Jing Wenjun)

- rgw: a few cases where rgw\_obj is incorrectly initialized ([issue#19146](#), [issue#19096](#), [pr#13843](#), Yehuda Sadeh)
- rgw: anonymous user error code of getting object is not consistent with SWIFT ([issue#18806](#), [issue#19178](#), [pr#13877](#), Jing Wenjun)
- rgw: civetweb frontend segfaults in Luminous ([issue#19749](#), [issue#19840](#), [pr#16166](#), Abhishek Lekshmanan, Jesse Williamson)
- rgw: civetweb: move to post 1.8 version ([issue#19704](#), [pr#14960](#), Yehuda Sadeh)
- rgw: "cluster [WRN] bad locator @X on object @X...." in cluster log ([issue#19212](#), [issue#18980](#), [pr#14065](#), Casey Bodley)
- rgw: crash when updating period with placement group ([issue#18772](#), [issue#18631](#), [pr#14511](#), Orit Wasserman)
- rgw: Custom data header support ([issue#19843](#), [pr#15985](#), Pavan Rallabhandi)
- rgw: datalog trim can't work as expected ([issue#20263](#), [issue#20190](#), [pr#16175](#), Zhang Shaowen)
- rgw: DUMPABLE flag is cleared by setuid preventing core dumps ([issue#19147](#), [issue#19089](#), [pr#13845](#), Brad Hubbard)
- rgw: Error parsing xml when get bucket lifecycle ([issue#19363](#), [issue#19534](#), [pr#14528](#), liuchang0812)
- rgw: first write also tries to read object ([issue#18904](#), [issue#18622](#), [issue#18623](#), [issue#18621](#), [pr#14515](#), Yehuda Sadeh)
- rgw: fix break inside of yield in RGWFetchAllMetaCR ([issue#19322](#), [issue#17655](#), [pr#14067](#), Casey Bodley)
- rgw: fix handling RGWUserInfo::system in RGWHandler\_REST\_SWIFT ([issue#18476](#), [pr#13006](#), Radoslaw Zarzynski)
- rgw: fix RadosGW hang during multi-chunk upload of AWSv4 ([issue#19837](#), [issue#19754](#), [pr#14939](#), Radoslaw Zarzynski)
- rgw: fix use of marker in List::list\_objects() ([issue#19047](#), [issue#18331](#), [pr#14517](#), Yehuda Sadeh)
- rgw: 'gc list -include-all' command infinite loop the first 1000 items ([issue#19978](#), [issue#20147](#), [pr#16139](#), Shasha Lu, fang yuxiang)
- rgw: get wrong content when download object with specific range when compression was enabled ([issue#20100](#), [issue#20268](#), [pr#16178](#), fang yuxiang)
- rgw: health check errors out incorrectly ([issue#19025](#), [issue#19157](#), [pr#13866](#), Pavan Rallabhandi)

- rgw: Lifecycle thread will still handle the bucket even if it has been removed ([issue#20285](#), [issue#20405](#), [pr#16183](#), Zhang Shaowen)
- rgw: make sending Content-Length in 204 and 304 controllable ([issue#18985](#), [issue#16602](#), [pr#13514](#), Radoslaw Zarzynski)
- rgw: meta sync thread crash at RGWMetaSyncShardCR ([issue#20251](#), [issue#20347](#), [pr#16180](#), Fang Yuxiang, Nathan Cutler)
- rgw: multisite: after CreateBucket is forwarded to master, local bucket may use different value for bucket index shards ([issue#19745](#), [issue#19759](#), [pr#16290](#), Shasha Lu)
- rgw: multisite: EPERM when trying to read SLO objects as system/admin user ([issue#19027](#), [issue#19475](#), [pr#14523](#), Casey Bodley)
- rgw: multisite: fetch\_remote\_obj() gets wrong version when copying from remote ([issue#19608](#), [pr#14606](#), Zhang Shaowen, Casey Bodley)
- rgw: multisite: RGWMetaSyncShardControlCR gives up on EIO ([issue#19160](#), [issue#19019](#), [pr#13868](#), Casey Bodley)
- rgw: multisite: segfault after changing value of rgw\_data\_log\_num\_shards ([issue#18488](#), [issue#18548](#), [pr#13181](#), Casey Bodley)
- rgw: multisite: some 'radosgw-admin data sync' commands hang ([issue#19236](#), [issue#19354](#), [pr#14142](#), Shasha Lu)
- rgw: multisite: some yields in RGWMetaSyncShardCR::full\_sync() resume in incremental\_sync() ([issue#19049](#), [issue#18076](#), [pr#13838](#), Casey Bodley)
- rgw: multisite: sync status reports master is on a different period ([issue#18709](#), [issue#18064](#), [pr#13176](#), Abhishek Lekshmanan)
- rgw: no http referer info in container metadata dump in swift API ([issue#18665](#), [issue#18898](#), [pr#13829](#), Jing Wenjun)
- rgw: "period update" does not remove short\_zone\_ids of deleted zones ([issue#15618](#), [issue#19342](#), [pr#14141](#), Casey Bodley)
- rgw: radosgw-admin: add the 'object stat' command to usage ([issue#19164](#), [issue#19013](#), [pr#13873](#), Pavan Rallabhandi)
- rgw: radosgw-admin period update reverts deleted zonegroup ([issue#18713](#), [issue#17239](#), [pr#13172](#), Orit Wasserman)
- rgw: 'radosgw-admin usage show' listing 0 bytes\_sent/received ([issue#20261](#), [pr#16174](#), Pritha Srivastava)
- rgw: 'radosgw-admin zone create' command with specified zone-id creates a zone with different id ([issue#19524](#), [issue#19498](#), [pr#14526](#), Orit Wasserman)

- rgw: Realm set does not create a new period ([issue#18333](#), [issue#18499](#), [pr#14509](#), Orit Wasserman)
- rgw: reduce log level of 'storing entry at' in cls\_log ([issue#19835](#), [issue#19839](#), [pr#16165](#), Willem Jan Withagen)
- rgw: Response header of swift API returned by radosgw does not contain x-openstack-request-id. But Swift returns it ([issue#19443](#), [issue#19573](#), [pr#14529](#), tone-zhang)
- rgw: rgw\_file: fix marker computation ([issue#20158](#), [issue#19526](#), [issue#18989](#), [issue#19470](#), [issue#19471](#), [issue#18651](#), [issue#20195](#), [issue#19059](#), [issue#19112](#), [issue#19018](#), [issue#19036](#), [issue#19154](#), [issue#19170](#), [issue#19663](#), [issue#19661](#), [issue#19111](#), [issue#18992](#), [issue#18650](#), [issue#18991](#), [issue#19623](#), [issue#19149](#), [issue#19270](#), [issue#19723](#), [issue#19625](#), [issue#19624](#), [issue#19060](#), [issue#19166](#), [issue#18810](#), [issue#19168](#), [issue#19162](#), [issue#19066](#), [issue#18808](#), [issue#19634](#), [issue#19435](#), [issue#19144](#), [issue#19229](#), [issue#18902](#), [pr#13871](#), Gui Hecheng, Matt Benjamin)
- rgw: S3 create bucket should not do response in json ([issue#19172](#), [issue#18889](#), [pr#13875](#), Abhishek Lekshmanan)
- rgw: S3 v4 authentication issue with X-Amz-Expires ([issue#19477](#), [issue#18828](#), [pr#14524](#), liuchang0812)
- rgw: S3 v4 authentication issue with X-Amz-Expires ([issue#19725](#), [issue#18828](#), [pr#16162](#), liuchang0812)
- rgw: should parse the url to http host to compare with the container referer acl ([issue#18896](#), [issue#18685](#), [pr#13780](#), Jing Wenjun)
- rgw: slave zonegroup cannot enable the bucket versioning ([issue#18711](#), [issue#18003](#), [pr#13174](#), Orit Wasserman)
- rgw: Swift API: spurious newline after http body causes weird errors ([issue#18780](#), [issue#18473](#), [pr#13224](#), Marcus Watts, Matt Benjamin)
- rgw: swift API: cannot disable object versioning with empty X-Versions-Location ([issue#18852](#), [issue#19175](#), [pr#14519](#), Jing Wenjun)
- rgw: swift: disable revocation thread under certain circumstances ([issue#19499](#), [issue#9493](#), [issue#19777](#), [pr#16164](#), Marcus Watts)
- rgw: Swift's at-root features (/crossdomain.xml, /info, /healthcheck) are broken ([issue#20031](#), [issue#19520](#), [pr#16168](#), Radoslaw Zarzynski)
- rgw: the swift container acl does not support field .ref ([issue#18909](#), [issue#19180](#), [issue#18484](#), [issue#18796](#), [pr#14516](#), Jing Wenjun, Radoslaw Zarzynski)
- rgw: typo in rgw\_admin.cc ([issue#19156](#), [issue#19026](#), [pr#13864](#), Ronak Jain)

- rgw: unsafe access in RGWListBucket\_ObjStore\_SWIFT::send\_response() ([issue#19574](#), [issue#19249](#), [pr#14530](#), Yehuda Sadeh)
- rgw: upgrade to multisite v2 fails if there is a zone without zone info ([issue#19331](#), [issue#19231](#), [pr#14137](#), Danny Al-Gaaf, Orit Wasserman)
- rgw: usage stats and quota are not operational for multi-tenant users ([issue#18364](#), [issue#18843](#), [issue#16355](#), [pr#14513](#), Radoslaw Zarzynski)
- rgw: Use decoded URI when verifying TempURL ([issue#18590](#), [issue#18627](#), [pr#12986](#), Michal Koutný)
- rgw: VersionIdMarker and NextVersionIdMarker are not returned when listing object versions ([issue#20363](#), [issue#19886](#), [pr#16181](#), Zhang Shaowen)
- rgw: when converting region\_map we need to use rgw\_zone\_root\_pool ([issue#19195](#), [issue#19356](#), [pr#14144](#), Orit Wasserman)
- rgw: when uploading the objects continuously in the versioned bucket, some objects will not sync ([issue#19766](#), [issue#18208](#), [pr#16163](#), lvshuhua)
- rgw: wrong object size after copy of uncompressed multipart objects ([issue#20269](#), [issue#20071](#), [pr#16179](#), fang yuxiang)
- rgw: zonegroupmap set does not work ([issue#18725](#), [issue#19479](#), [pr#14525](#), Casey Bodley)
- tests: AttributeError: Thrasher instance has no attribute 'ceph\_objectstore\_tool' ([issue#19064](#), [issue#18799](#), [pr#13609](#), Nathan Cutler)
- tests: backport Sage's fixes to qa/suites/upgrade/jewel-x ([issue#19651](#), [pr#14612](#), Sage Weil)
- tests: ceph-object-corpus: kraken objects ([issue#20878](#), [pr#14983](#), Sage Weil)
- tests: CMakeLists.txt: disable memstore make check test ([issue#17743](#), [pr#16215](#), Sage Weil)
- tests: HEALTH\_WARN pool rbd pg\_num 244 > pgp\_num 224 during upgrade ([issue#19771](#), [issue#20024](#), [pr#16137](#), Kefu Chai)
- tests: ignore bogus ceph-objectstore-tool error in ceph\_manager ([issue#18805](#), [issue#16263](#), [pr#13239](#), Nathan Cutler, Kefu Chai)
- tests: insufficient timeout in radosbench task ([issue#20497](#), [pr#16111](#), Sage Weil)
- tests: LibRadosMiscConnectFailure.ConnectFailure hang ([issue#20271](#), [issue#19901](#), [pr#16140](#), Sage Weil)
- tests: [librados\_test\_stub] cls\_cxx\_map\_get\_XYZ methods don't return correct value ([issue#19597](#), [issue#19609](#), [pr#16097](#), Jason Dillaman)

- tests: move swift.py task from teuthology to ceph, phase one (kraken) ([issue#20392](#), [pr#15869](#), Nathan Cutler, Sage Weil, Warren Usui, Greg Farnum, Ali Maredia, Tommi Virtanen, Zack Cerza, Sam Lang, Yehuda Sadeh, Joe Buck, Josh Durgin)
- tests: ObjectStore/StoreTest.OnodeSizeTracking/2 fails on bluestore ([issue#20499](#), [pr#16112](#), xie xingguo)
- tests: qa: ceph-ansible test tweaks ([issue#20882](#), [pr#12984](#), [pr#13618](#), Tamil Muthamizhan, Yuri Weinstein)
- tests: qa/suites/upgrade: add tiering test to hammer-jewel-x ([issue#20879](#), [issue#19185](#), [pr#14692](#), Kefu Chai)
- tests: qa/tasks: misc systemd updates ([issue#19719](#), [pr#14702](#), Vasu Kulkarni)
- tests: qa/tasks: rbd-mirror daemon not properly run in foreground mode ([issue#20638](#), [issue#20630](#), [issue#20634](#), [pr#16342](#), Jason Dillaman)
- tests: qa/tasks: set pgp = pg num on thrashing finish ([issue#20881](#), [pr#13757](#), Kefu Chai)
- tests: qa/tasks/workunit: Backport repo fixes from master ([issue#19429](#), [issue#19531](#), [pr#14487](#), Kefu Chai, Dan Mick)
- tests: remove hard-coded image name from TestLibRBD.Mirror ([issue#18555](#), [issue#19130](#), [issue#19227](#), [issue#18447](#), [issue#19807](#), [issue#19798](#), [pr#16113](#), Mykola Golub, Jason Dillaman)
- tests: remove qa/suites/buildpackages ([issue#18849](#), [issue#18846](#), [pr#13298](#), Loic Dachary)
- tests: run certain upgrade/jewel-x tests on Xenial only ([issue#20877](#), [pr#16493](#), Nathan Cutler)
- tests: run-rbd-unit-tests.sh assert in lockdep\_will\_lock, TestLibRBD.ObjectMapConsistentSnap ([issue#18822](#), [issue#17447](#), [pr#14151](#), Jason Dillaman)
- tests: SUSE yaml facets in qa/distros/all are out of date ([issue#18849](#), [issue#18870](#), [issue#18846](#), [issue#18856](#), [pr#13330](#), Nathan Cutler)
- tests: swift.py: clone the ceph-kraken branch ([issue#20520](#), [pr#16131](#), Nathan Cutler)
- tests: test/librbd: decouple ceph\_test\_librbd\_api from libceph-common ([issue#20175](#), [issue#20351](#), [pr#16195](#), Kefu Chai)
- tests: test\_notify.py: assert(not image.is\_exclusive\_lock\_owner()) on line 147 ([issue#19716](#), [issue#19794](#), [pr#14833](#), Mykola Golub)

- tests: test\_notify.py: rbd.InvalidArgument: error updating features for image test\_notify\_clone2 ([issue#19692](#), [issue#19693](#), [pr#14641](#), Jason Dillaman)
- tests: use ceph-kraken branch for s3tests ([issue#18387](#), [pr#12746](#), Nathan Cutler)
- tests: use librados API to retrieve config params ([issue#18668](#), [issue#18617](#), [pr#13102](#), Jason Dillaman)
- tests: various OpenStack tweaks ([issue#20882](#), [pr#13707](#), [pr#13641](#), [pr#13635](#), [pr#13633](#), [pr#13613](#), [pr#13283](#), [pr#13673](#), [pr#13638](#), [pr#14485](#), Zack Cerza)
- tools: ceph-brag fails to count “in” mds ([issue#19333](#), [issue#19192](#), [pr#14098](#), Peng Zhang)
- tools: ceph-disk prepare writes osd log 0 with root owner ([issue#18538](#), [issue#18606](#), [pr#13026](#), Samuel Matzek)
- tools: RadosImport::import should return an error if Rados::connect fails ([issue#19351](#), [issue#19319](#), [pr#14095](#), Brad Hubbard)

## v11.2.0 Kraken

---

This is the first release of the Kraken series. It is a stable release that will be maintained with bugfixes and backports until the next stable release, Luminous, is completed in the Spring of 2017.

## Major Changes from Jewel

---

- *RADOS:*
  - The new *BlueStore* backend now has a stable disk format and is passing our failure and stress testing. Although the backend is still flagged as experimental, we encourage users to try it out for non-production clusters and non-critical data sets.
  - RADOS now has experimental support for *overwrites on erasure-coded pools*. Because the disk format and implementation are not yet finalized, there is a special pool option that must be enabled to test the new feature. Enabling this option on a cluster will permanently bar that cluster from being upgraded to future versions.
  - We now default to the *AsyncMessenger* (`ms type = async`) instead of the legacy *SimpleMessenger*. The most noticeable difference is that we now use a fixed sized thread pool for network connections (instead of two threads per socket with *SimpleMessenger*).
  - Some OSD failures are now detected almost immediately, whereas previously the heartbeat timeout (which defaults to 20 seconds) had to expire. This prevents

IO from blocking for an extended period for failures where the host remains up but the ceph-osd process is no longer running.

- There is a new `ceph-mgr` daemon. It is currently collocated with the monitors by default, and is not yet used for much, but the basic infrastructure is now in place.
  - The size of encoded OSDMaps has been reduced.
  - The OSDs now quiesce scrubbing when recovery or rebalancing is in progress.
- *RGW:*
    - RGW now supports a new zone type that can be used for metadata indexing via ElasticSearch.
    - RGW now supports the S3 multipart object copy-part API.
    - It is possible now to reshuffle an existing bucket. Note that bucket reshuffling currently requires that all IO (especially writes) to the specific bucket is quiesced.
    - RGW now supports data compression for objects.
    - Civetweb version has been upgraded to 1.8
    - The Swift static website API is now supported (S3 support has been added previously).
    - S3 bucket lifecycle API has been added. Note that currently it only supports object expiration.
    - Support for custom search filters has been added to the LDAP auth implementation.
    - Support for NFS version 3 has been added to the RGW NFS gateway.
    - A Python binding has been created for `librgw`.
  - *RBD:*
    - RBD now supports images stored in an *erasure-coded* RADOS pool using the new (experimental) overwrite support. Images must be created using the new `rbd` CLI “`-data-pool <ec pool>`” option to specify the EC pool where the backing data objects are stored. Attempting to create an image directly on an EC pool will not be successful since the image’s backing metadata is only supported on a replicated pool.
    - The `rbd-mirror` daemon now supports replicating dynamic image feature updates and image metadata key/value pairs from the primary image to the non-primary image.

- The number of image snapshots can be optionally restricted to a configurable maximum.
  - The rbd Python API now supports asynchronous IO operations.
- *CephFS*:
    - libcephfs function definitions have been changed to enable proper uid/gid control. The library version has been increased to reflect the interface change.
    - Standby replay MDS daemons now consume less memory on workloads doing deletions.
    - Scrub now repairs backtrace, and populates damage ls with discovered errors.
    - A new pg\_files subcommand to cephfs-data-scan can identify files affected by a damaged or lost RADOS PG.
    - The false-positive “failing to respond to cache pressure” warnings have been fixed.

## Upgrading from Kraken release candidate 11.1.0

---

- The new *BlueStore* backend had an on-disk format change after 11.1.0. Any BlueStore OSDs created with 11.1.0 will need to be destroyed and recreated.

## Upgrading from Jewel

- All clusters must first be upgraded to Jewel 10.2.z before upgrading to Kraken 11.2.z (or, eventually, Luminous 12.2.z).
- The `sortbitwise` flag must be set on the Jewel cluster before upgrading to Kraken. The latest Jewel (10.2.8+) releases issue a health warning if the flag is not set, so this is probably already set. If it is not, Kraken OSDs will refuse to start and will print an error message in their log.
- You may upgrade OSDs, Monitors, and MDSs in any order. RGW daemons should be upgraded last.
- When upgrading, new ceph-mgr daemon instances will be created automatically alongside any monitors. This will be true for Jewel to Kraken and Jewel to Luminous upgrades, but likely not be true for future upgrades beyond Luminous. You are, of course, free to create new ceph-mgr daemon instances and destroy the auto-created ones if you do not want them to be colocated with the ceph-mon daemons.

## BlueStore

BlueStore is a new backend for managing data stored by each OSD on the directly hard disk or SSD. Unlike the existing FileStore implementation, which makes use of an XFS file system to store objects as files, BlueStore manages the underlying block device directly. Implements its own file system-like on-disk structure designed specifically for Ceph OSD workloads. Key features of BlueStore include:

- Checksums on all data written to disk, with checksum verifications on all reads, enabled by default.
- Inline compression support, which can be enabled on a per-pool or per-object basis via pool properties or client hints, respectively.
- Efficient journaling. Unlike FileStore, which writes *all* data to its journal device, BlueStore only journals metadata and (in some cases) small writes, reducing the size and throughput requirements for its journal. As with FileStore, the journal can be colocated on the same device as other data or allocated on a smaller, high-performance device (e.g., an SSD or NVMe device). BlueStore journals are only 512 MB by default.

The BlueStore on-disk format is expected to continue to evolve. However, we will provide support in the OSD to migrate to the new format on upgrade.

In order to enable BlueStore, add the following to `ceph.conf`:

```
1. enable experimental unrecoverable data corrupting features = bluestore
```

To create a BlueStore OSD, pass the `-bluestore` option to `ceph-disk` or `ceph-deploy`

during OSD creation.

## Upgrade notes

- The OSDs now avoid starting new scrubs while recovery is in progress. To revert to the old behavior (and do not let recovery activity affect the scrub scheduling) you can set the following option:

```
1. osd scrub during recovery = true
```

- The list of monitor hosts/addresses for building the monmap can now be obtained from DNS SRV records. The service name used when querying the DNS is defined in the “mon\_dns\_srv\_name” config option, which defaults to “ceph-mon”.
- The ‘osd class load list’ config option is a list of object class names that the OSD is permitted to load (or ‘\*’ for all classes). By default it contains all existing in-tree classes for backwards compatibility.
- The ‘osd class default list’ config option is a list of object class names (or ‘\*’ for all classes) that clients may invoke having only the ‘\*’, ‘x’, ‘class-read’, or ‘class-write’ capabilities. By default it contains all existing in-tree classes for backwards compatibility. Invoking classes not listed in ‘osd class default list’ requires a capability naming the class (e.g. ‘allow class foo’).
- The ‘rgw rest getusage op compat’ config option allows you to dump (or not dump) the description of user stats in the S3 GetUsage API. This option defaults to false. If the value is true, the response data for GetUsage looks like:

```
1. "stats": {
2.     "TotalBytes": 516,
3.     "TotalBytesRounded": 1024,
4.     "TotalEntries": 1
5. }
```

If the value is false, the response for GetUsage looks as it did before:

```
1. {
2.     516,
3.     1024,
4.     1
5. }
```

- The ‘osd out ...’ and ‘osd in ...’ commands now preserve the OSD weight. That is, after marking an OSD out and then in, the weight will be the same as before (instead of being reset to 1.0). Previously the mons would only preserve the weight if the mon automatically marked and OSD out and then in, but not when an admin did so explicitly.

- The ‘ceph osd perf’ command will display ‘commit\_latency(ms)’ and ‘apply\_latency(ms)’. Previously, the names of these two columns are ‘fs\_commit\_latency(ms)’ and ‘fs\_apply\_latency(ms)’. We remove the prefix ‘fs\_’, because they are not filestore specific.
- Monitors will no longer allow pools to be removed by default. The setting `mon_allow_pool_delete` has to be set to true (defaults to false) before they allow pools to be removed. This is a additional safeguard against pools being removed by accident.
- If you have manually specified the monitor user `rocksdb` via the `mon keyvaluedb = rocksdb` option, you will need to manually add a file to the mon data directory to preserve this option:

```
1. echo rocksdb > /var/lib/ceph/mon/ceph-`hostname`/kv_backend
```

New monitors will now use `rocksdb` by default, but if that file is not present, existing monitors will use `leveldb`. The `mon keyvaluedb` option now only affects the backend chosen when a monitor is created.

- The ‘osd crush initial weight’ option allows you to specify a CRUSH weight for a newly added OSD. Previously a value of 0 (the default) meant that we should use the size of the OSD’s store to weight the new OSD. Now, a value of 0 means it should have a weight of 0, and a negative value (the new default) means we should automatically weight the OSD based on its size. If your configuration file explicitly specifies a value of 0 for this option you will need to change it to a negative value (e.g., -1) to preserve the current behavior.
- The `osd crush location` config option is no longer supported. Please update your `ceph.conf` to use the `crush location` option instead.
- The static libraries are no longer included by the debian development packages (`lib*-dev`) as it is not required per debian packaging policy. The shared (.so) versions are packaged as before.
- The libtool pseudo-libraries (.la files) are no longer included by the debian development packages (`lib*-dev`) as they are not required per <https://wiki.debian.org/ReleaseGoals/LAFileRemoval> and <https://www.debian.org/doc/manuals/maint-guide/advanced.en.html>.
- The jerasure and shc plugins can now detect SIMD instruction at runtime and no longer need to be explicitly configured for different processors. The following plugins are now deprecated: `jerasure_generic`, `jerasure_sse3`, `jerasure_sse4`, `jerasure_neon`, `shc_generic`, `shc_sse3`, `shc_sse4`, and `shc_neon`. If you use any of these plugins directly you will see a warning in the mon log file. Please switch to using just ‘`jerasure`’ or ‘`shc`’.
- The librados omap `get_keys` and `get_vals` operations include a start key and a

limit on the number of keys to return. The OSD now imposes a configurable limit on the number of keys and number of total bytes it will respond with, which means that a librados user might get fewer keys than they asked for. This is necessary to prevent careless users from requesting an unreasonable amount of data from the cluster in a single operation. The new limits are configured with

`osd_max omap_entries_per_request`, defaulting to 131,072, and `osd_max omap_bytes_per_request`, defaulting to 4MB.

- Calculation of recovery priorities has been updated. This could lead to unintuitive recovery prioritization during cluster upgrade. In case of such recovery, OSDs in the old version would operate on different priority ranges than new ones. Once upgraded, the cluster will operate on consistent values.

# Notable Changes

- bluestore: add counter to trace blob splitting ([pr#11718](#), xie xingguo)
- bluestore: a few more cleanups ([pr#11780](#), xie xingguo)
- bluestore: avoid polluting shard info if need resharding ([pr#11439](#), xie xingguo)
- bluestore: avoid unnecessary call to init\_csum() ([pr#12015](#), xie xingguo)
- bluestore: ceph-disk: adjust bluestore default device sizes ([pr#12530](#), Sage Weil)
- bluestore: ceph\_test\_objectstore: smaller device ([pr#11591](#), Sage Weil)
- bluestore: clean up Allocator::dump ([issue#18054](#), [pr#12282](#), Sage Weil)
- bluestore: clear extent map on object removal ([pr#11603](#), Sage Weil)
- bluestore: compressor/ZLibCompressor: fix broken isal-l ([pr#11445](#), Igor Fedotov)
- bluestore: dedup if space overlap truly exists ([pr#11986](#), xie xingguo)
- bluestore: dedup omap\_head, reuse nid instead ([pr#12275](#), xie xingguo)
- bluestore: deep fsck ([pr#11724](#), Sage Weil)
- bluestore: default bluestore\_clone\_cow=true ([pr#11540](#), Sage Weil)
- bluestore: drop inline\_dirty from struct ExtentMap ([pr#11377](#), xie xingguo)
- bluestore: drop member "space" from Onode ([pr#12185](#), xie xingguo)
- bluestore: fix alloc release timing on sync submits ([pr#11983](#), Sage Weil)
- bluestore: fix bufferspace stats leak due to blob splitting ([pr#12039](#), xie xingguo)
- bluestore: fix collection\_list end bound off-by-one ([pr#11771](#), Sage Weil)
- bluestore: fix compiler warnings ([pr#11905](#), xie xingguo)
- bluestore: fixes and cleanups ([pr#11761](#), xie xingguo)
- bluestore: fix escaping of chars > 0x80 ([pr#11502](#), Sage Weil)
- bluestore: fix extent shard span check ([pr#11725](#), Sage Weil)
- bluestore: fix has\_aios ([pr#11317](#), Sage Weil)
- bluestore: Fix invalid compression statfs caused by clone op ([pr#11351](#), Igor Fedotov)

- bluestore: fix lack of resharding ([pr#11597](#), Igor Fedotov)
- bluestore: fix latency calculation ([pr#12040](#), Pan Liu)
- bluestore: fix onode vs extent key suffix ([pr#11452](#), Sage Weil)
- bluestore: fix potential memory leak ([pr#11893](#), xie xingguo)
- bluestore: fix race condition during blob splitting ([pr#11422](#), xiexingguo, xie xingguo)
- bluestore: fix remove\_collection to properly detect collection e... ([pr#11398](#), Igor Fedotov)
- bluestore: fix \_split\_collections race with osr\_reap ([pr#11748](#), Sage Weil)
- bluestore: fix up compression tests and debug output ([pr#11350](#), Sage Weil)
- bluestore: fix writes that span existing shard boundaries ([pr#11451](#), Sage Weil)
- bluestore: flush before enumerating omap values ([issue#18140](#), [pr#12328](#), Sage Weil)
- bluestore: formatting nits ([pr#11514](#), xie xingguo)
- bluestore: fsck: fix omap\_head check ([pr#11726](#), Sage Weil)
- bluestore: GC infra refactor, more UTs and GC range calculation fixes ([pr#11482](#), Igor Fedotov)
- bluestore: KernelDevice: fix race in aio\_thread vs aio\_wait ([issue#17824](#), [pr#12204](#), Sage Weil)
- bluestore: kv: dump rocksdb stats ([pr#12287](#), Varada Kari, Jianpeng Ma, Sage Weil)
- bluestore: kv/rocksdb: enable rocksdb write path breakdown ([pr#11696](#), Haodong Tang)
- bluestore: kv/RocksDBStore: rename option ([pr#11769](#), Sage Weil)
- bluestore: less code redundancy ([pr#11740](#), xie xingguo)
- bluestore: make 2q cache kin/kout size tunable ([pr#11599](#), Haodong Tang)
- bluestore: mark ops that can't tolerate ENOENT ([pr#12114](#), Sage Weil)
- bluestore: mempool: changes for bitmap allocator ([pr#11922](#), Ramesh Chander)
- bluestore: misc. fixes and cleanups ([pr#11964](#), xie xingguo)
- bluestore: move bluefs into its own mempool ([pr#11834](#), Sage Weil)
- bluestore: no garbage collection for uncompressed blobs ([pr#11539](#), Roushan Ali,

Sage Weil)

- bluestore: optional debug mode to identify aio stalls ([pr#11818](#), Sage Weil)
- bluestore: os/bluestore: a few cleanups ([pr#11483](#), Sage Weil)
- bluestore: os/bluestore: avoid resharding if the last shard size fall below shar... ([pr#12447](#), Igor Fedotov)
- bluestore: os/bluestore: bitmap allocator dump functionality ([pr#12298](#), Ramesh Chander)
- bluestore: os/bluestore: bluestore\_sync\_submit\_transaction = false ([pr#12367](#), Sage Weil)
- bluestore: os/bluestore: cleanup around Blob::ref\_map ([pr#11896](#), Igor Fedotov)
- bluestore: os/bluestore: clear omap flag if parent has none ([pr#12351](#), xie xingguo)
- bluestore: os/bluestore: don't implicitly create the source object for clone ([pr#12353](#), xie xingguo)
- bluestore: os/bluestore: drop old bluestore preconditioning; replace with wal preextension of file size ([pr#12265](#), Sage Weil)
- bluestore: os/bluestore: fix global commit latency ([pr#12356](#), xie xingguo)
- bluestore: os/bluestore: fix ondisk encoding for blobs ([pr#12488](#), Varada Kari, Sage Weil)
- bluestore: os/bluestore: fix potential csum\_order overflow ([pr#12333](#), xie xingguo)
- bluestore: os/bluestore: fix target\_buffer value overflow in Cache::trim() ([pr#12507](#), Igor Fedotov)
- bluestore: os/bluestore: include modified objects in flush list even if onode unchanged ([pr#12541](#), Sage Weil)
- bluestore: os/bluestore: kill dead gc-related counters ([pr#12065](#), xie xingguo)
- bluestore: os/bluestore: kill overlay related options ([pr#11557](#), xie xingguo)
- bluestore: os/bluestore: misc coverity fixes/cleanups ([pr#12202](#), Sage Weil)
- bluestore: os/bluestore: preserve source collection cache during split ([pr#12574](#), Sage Weil)
- bluestore: os/bluestore: remove 'extents' from shard\_info ([pr#12629](#), Sage Weil)
- bluestore: os/bluestore: simplified allocator interfaces to single apis

([pr#12355](#), Ramesh Chander)

- bluestore: os/bluestore: simplify allocator release flow ([pr#12343](#), Sage Weil)
- bluestore: os/bluestore: simplify can\_split\_at() ([pr#11607](#), xie xingguo)
- bluestore: os/bluestore: use iterator for erase() method directly ([pr#11490](#), xie xingguo)
- bluestore: os/kstore: rmcoll fix to satisfy store\_test ([pr#11533](#), Igor Fedotov)
- bluestore: os: make filestore\_blackhole -> objectstore\_blackhole ([pr#11788](#), Sage Weil)
- bluestore: os: move\_ranges\_destroy\_src ([pr#11237](#), Manali Kulkarni, Sage Weil)
- bluestore: readability improvements and doxygen fix ([pr#11895](#), xie xingguo)
- bluestore: reap collection after all pending ios done ([pr#11797](#), Haomai Wang)
- bluestore: reap ioc when stopping aio\_thread. ([pr#11811](#), Haodong Tang)
- bluestore: refactor \_do\_write(); move initializaiton of csum out of loop ([pr#11823](#), xie xingguo)
- bluestore: remove duplicated namespace of tx state ([pr#11845](#), xie xingguo)
- bluestore: remove garbage collector staff ([pr#12042](#), Igor Fedotov)
- bluestore: set next object as gobject\_t::get\_max() when start.hobj.i... ([pr#11495](#), Xinze Chi, Haomai Wang)
- bluestore: simplify blob status checking for small writes ([pr#11366](#), xie xingguo)
- bluestore: some more cleanups ([pr#11910](#), xie xingguo)
- bluestore: spdk: a few fixes ([pr#11882](#), Yehuda Sadeh)
- bluestore: speed up omap-key generation for same onode ([pr#11807](#), xie xingguo)
- bluestore: traverse buffer\_map in reverse order when splitting BufferSpace ([pr#11468](#), xie xingguo)
- bluestore: update cache logger after 'trim\_cache' operation ([pr#11695](#), Haodong Tang)
- bluestore: use bitmap allocator for bluefs ([pr#12285](#), Sage Weil)
- bluestore: use std::unordered\_map for SharedBlob lookup ([pr#11394](#), Sage Weil)
- build/ops: AArch64: Detect crc32 extension support from assembler ([issue#17516](#), [pr#11391](#), Alexander Graf)

- build/ops: boost: embedded ([pr#11817](#), Sage Weil, Matt Benjamin)
- build/ops: build: dump env during build ([issue#18084](#), [pr#12284](#), Sage Weil)
- build/ops: ceph-detect-init: FreeBSD introduction of bsdrc ([pr#11906](#), Willem Jan Withagen, Kefu Chai)
- build/ops: ceph-disk: enable -runtime ceph-osd systemd units ([issue#17889](#), [pr#12241](#), Loic Dachary)
- build/ops: ceph.spec: add pybind rgwfile ([pr#11847](#), Haomai Wang)
- build/ops,cleanup,bluestore: os/bluestore: remove build warning in a better way ([pr#11920](#), Igor Fedotov)
- build/ops: CMakeLists: add vstart-base target ([pr#12476](#), Sage Weil)
- build/ops: CMakeLists.txt: enable LTTNG by default ([pr#11500](#), Sage Weil)
- build/ops: common/buffer.cc: raw\_pipe depends on splice(2) ([pr#11967](#), Willem Jan Withagen)
- build/ops,common: common/str\_list.h: fix clang warning about std::move ([pr#12570](#), Willem Jan Withagen)
- build/ops,core: xio: fix build ([pr#11768](#), Matt Benjamin)
- build/ops: deb: add python dependencies where needed ([issue#17579](#), [pr#11507](#), Nathan Cutler, Kefu Chai)
- build/ops: deb: add python-rgw packages ([pr#11832](#), Sage Weil)
- build/ops: debian: apply dh\_python to python-rgw also ([pr#12260](#), Kefu Chai)
- build/ops: deb: update python-rgw dependencies to librgw2 ([pr#11885](#), Casey Bodley)
- build/ops: do\_freebsd.sh: Build with SYSTEM Boost on FreeBSD ([pr#11942](#), Willem Jan Withagen)
- build/ops: do\_freebsd.sh: Do not use LTTNG on FreeBSD ([pr#11551](#), Willem Jan Withagen)
- build/ops: do\_freebsd.sh: Set options for debug building. ([pr#11443](#), Willem Jan Withagen)
- build/ops: FreeBSD: do\_freebsd.sh ([pr#12090](#), Willem Jan Withagen)
- build/ops: FreeBSD: test/encoding/readable.sh": fix nproc and ls -v calls ([pr#11522](#), Willem Jan Withagen)
- build/ops: FreeBSD: update require packages ([pr#11512](#), Willem Jan Withagen)

- build/ops: git-archive-all.sh: use an actually unique tmp dir ([pr#12011](#), Dan Mick)
- build/ops: include/enc: make clang happy ([pr#11638](#), Kefu Chai, Sage Weil)
- build/ops: install-deps.sh: allow building on SLES systems ([pr#11708](#), Nitin A Kamble)
- build/ops: install-deps.sh: JQ is needed in one script ([pr#12080](#), Willem Jan Withagen)
- build/ops: Log: Replace namespace log with logging ([pr#11650](#), Willem Jan Withagen)
- build/ops: Merging before make check because it clearly breaks the build and the build part is done ([pr#11924](#), Sage Weil)
- build/ops: ok, w/upstream acks, merging-jenkins build did succeed (this is a build-only change) ([pr#12008](#), Matt Benjamin)
- build/ops: qa: Add ceph-ansible installer. ([issue#16770](#), [pr#10402](#), Warren Usui)
- build/ops: rocksdb: do not build with -march=native ([pr#11677](#), Kefu Chai)
- build/ops: rocksdb: update to latest ([pr#12100](#), Kefu Chai)
- build/ops: rpm: Remove trailing whitespace in usermod command (SUSE) ([pr#10707](#), Tim Serong)
- build/ops: scripts/release-notes: allow title guesses from gh tags & description update ([pr#11399](#), Abhishek Lekshmanan)
- build/ops: systemd: Fix startup of ceph-mgr on Debian 8 ([pr#12555](#), Mark Korenberg)
- build/ops: tracing/objectstore.tp: add missing move\_ranges\_... tp ([pr#11484](#), Sage Weil)
- build/ops: upstart: fix ceph-crush-location default ([issue#6698](#), [pr#803](#), Jason Dillaman)
- build/ops: upstart: start ceph-all after static-network-up ([issue#17689](#), [pr#11631](#), Billy Olsen)
- cephfs: add gid to asok status ([pr#11487](#), Patrick Donnelly)
- cephfs: API cleanup for libcephfs interfaces ([issue#17911](#), [pr#12106](#), Jeff Layton)
- cephfs: ceph-fuse: start up log on parent process before shutdown ([issue#18157](#), [pr#12347](#), Greg Farnum)
- cephfs: ceph\_fuse: use sizeof get the buf length ([pr#11176](#), LeoZhang)

- cephfs, cleanup: ceph-fuse: start up log on parent process before shutdown ([issue#18157](#), [pr#12358](#), Kefu Chai)
- cephfs: client: add pid to metadata ([issue#17276](#), [pr#11359](#), Patrick Donnelly)
- cephfs: client: Client.cc: remove duplicated op type checking against CEPH\_MD... ([pr#11608](#), Weibing Zhang)
- cephfs: client: don't take extra target inode reference in ll\_link ([pr#11440](#), Jeff Layton)
- cephfs: client: fix mutex name typos ([pr#12401](#), Yunchuan Wen)
- cephfs: client: get caller's uid/gid on every libcephfs operation ([issue#17591](#), [pr#11526](#), Yan, Zheng)
- cephfs: client: get gid from MonClient ([pr#11486](#), Patrick Donnelly)
- cephfs: client: improve failure messages/debugging ([pr#12110](#), Patrick Donnelly)
- cephfs: client/mds: Clear setuid bits when writing or truncating ([issue#18131](#), [pr#12412](#), Jeff Layton)
- cephfs: client: put CapSnap not ptr in cap\_snaps map ([pr#12111](#), Patrick Donnelly)
- cephfs: client: remove redundant initialization ([pr#12028](#), Patrick Donnelly)
- cephfs: client: remove unnecessary bufferptr[] for writev ([pr#11836](#), Patrick Donnelly)
- cephfs: client: remove unneeded layout on MClientCaps ([pr#11790](#), John Spray)
- cephfs: client: set metadata["root"] from mount method when it's called with ... ([pr#12505](#), Jeff Layton)
- cephfs: client: trim\_caps() do not dereference cap if it's removed ([pr#12145](#), Kefu Chai)
- cephfs: client: use unique\_ptr ([pr#11837](#), Patrick Donnelly)
- cephfs: common/ceph\_string: add ceph string constants for CEPH\_SESSION\_FORCE\_RO ([pr#11516](#), Zhi Zhang)
- cephfs: Fix #17562 (backtrace check fails when scrubbing directory created by fsstress) ([issue#17562](#), [pr#11517](#), Yan, Zheng)
- cephfs: fix missing ll\_get for ll\_walk ([issue#18086](#), [pr#12061](#), Gui Hecheng)
- cephfs: get new fsmap after marking clusters down ([issue#7271](#), [issue#17894](#), [pr#1262](#), Patrick Donnelly)
- cephfs: Have ceph clear setuid/setgid bits on chown ([issue#18131](#), [pr#12331](#), Jeff

Layton)

- cephfs: libcephfs: add ceph\_fsetattr&&ceph\_lchmod&&ceph\_lutime ([pr#11191](#), huanwen ren)
- cephfs: libcephfs: add readlink function in cephfs.pyx ([pr#12384](#), huanwen ren)
- cephfs: libcephfs and test suite fixes ([issue#18013](#), [issue#17982](#), [pr#12228](#), Jeff Layton)
- cephfs: libcephfs client API overhaul and update ([pr#11647](#), Jeff Layton)
- cephfs: lua: use simpler lua\_next traversal structure ([pr#11958](#), Patrick Donnelly)
- cephfs: mds/Beacon: move C\_MDS\_BeaconSender class to .cc ([pr#10940](#), Michal Jarzabek)
- cephfs: mds/CDir.cc: remove unneeded use of count ([pr#11613](#), Michal Jarzabek)
- cephfs: mds/CINode.h: remove unneeded use of count ([pr#11371](#), Michal Jarzabek)
- cephfs: mds/DamageTable.cc: move shared ptrs ([pr#11435](#), Michal Jarzabek)
- cephfs: mds/DamageTable.cc: remove unneeded use of count ([pr#11625](#), Michal Jarzabek)
- cephfs: mds/DamageTable: move classes to .cc file ([pr#11450](#), Michal Jarzabek)
- cephfs: mds/flock: add const to member functions ([pr#11692](#), Michal Jarzabek)
- cephfs: mds/FSMap.cc: remove unneeded use of count ([pr#11402](#), Michal Jarzabek)
- cephfs: mds/FSMapUser.h: remove copy ctr and assign op ([pr#11509](#), Michal Jarzabek)
- cephfs: mds/InfoTable.h: add override to virtual functs ([pr#11496](#), Michal Jarzabek)
- cephfs: mds/InoTable.h: add override to virtual functs ([pr#11604](#), Michal Jarzabek)
- cephfs: mds/Mantle.h: include correct header files ([pr#11886](#), Michal Jarzabek)
- cephfs: mds/Mantle: pass parameters by const ref ([pr#11713](#), Michal Jarzabek)
- cephfs: mds/MDCache.h: remove unneeded call to clear func ([pr#11954](#), Michal Jarzabek)
- cephfs: mds/MDCache.h: remove unused functions ([pr#11908](#), Michal Jarzabek)
- cephfs: mds/MDLog: add const to member functions ([pr#11663](#), Michal Jarzabek)

- cephfs: mds/MDSMap.h: add const to member functions ([pr#11511](#), Michal Jarzabek)
- cephfs: mds/MDSRank: add const to member functions ([pr#11752](#), Michal Jarzabek)
- cephfs: mds/MDSRank.h: add override to virtual function ([pr#11727](#), Michal Jarzabek)
- cephfs: mds/MDSRank.h: make destructor protected ([pr#11651](#), Michal Jarzabek)
- cephfs: mds/MDSTableClient.h: add const to member funct ([pr#11681](#), Michal Jarzabek)
- cephfs: mds/Migrator.cc: remove unneeded use of count ([pr#11523](#), Michal Jarzabek)
- cephfs: mds/Migrator.h: add const to member functions ([pr#11819](#), Michal Jarzabek)
- cephfs: mds/Migrator.h: remove unneeded use of count ([pr#11833](#), Michal Jarzabek)
- cephfs: mds/Mutation.h: add const to member functions ([pr#11670](#), Michal Jarzabek)
- cephfs: mds/Mutation.h: simplify constructors ([pr#11455](#), Michal Jarzabek)
- cephfs: MDS: reduce usage of context wrapper ([pr#11560](#), Yan, Zheng)
- cephfs: mds/ScrubHeader.h: pass string by const reference ([pr#11904](#), Michal Jarzabek)
- cephfs: mds/server: merge the snapshot request judgment ([pr#11150](#), huanwen ren)
- cephfs: mds/SessionMap: add const to member functions ([pr#11541](#), Michal Jarzabek)
- cephfs: mds/SessionMap.cc: avoid copying and add const ([pr#11297](#), Michal Jarzabek)
- cephfs: mds/SessionMap.cc:put classes in unnamed namespace ([pr#11316](#), Michal Jarzabek)
- cephfs: mds/SessionMap.cc: remove unneeded use of count ([pr#11338](#), Michal Jarzabek)
- cephfs: mds/SessionMap.h: remove unneeded function ([pr#11565](#), Michal Jarzabek)
- cephfs: mds/SessionMap.h: remove unneeded use of count ([pr#11358](#), Michal Jarzabek)
- cephfs: mds/SnapRealm: remove unneeded use of count ([pr#11609](#), Michal Jarzabek)
- cephfs: mds/SnapServer.h: add override to virtual functs ([pr#11380](#), Michal Jarzabek)
- cephfs: mds/SnapServer.h: add override to virtual functs ([pr#11583](#), Michal Jarzabek)

- cephfs: mon/MDSMonitor: fix iterating over mutated map ([issue#18166](#), [pr#12395](#), John Spray)
- cephfs: multimds: fix state check in Migrator::find\_stale\_export\_freeze() ([pr#12098](#), Yan, Zheng)
- cephfs: osdc: After write try merge bh. ([issue#17270](#), [pr#11545](#), Jianpeng Ma)
- cephfs: Partial organization of mds/ header sections ([pr#11959](#), Patrick Donnelly)
- cephfs: Port/bootstrap ([pr#827](#), Yan, Zheng)
- cephfs: Revert “osdc: After write try merge bh.” ([issue#17270](#), [pr#11262](#), John Spray)
- cephfs: Small pile of random cephfs fixes and cleanup ([pr#11421](#), Jeff Layton)
- cephfs: src/mds: fix MDSMap upgrade decoding ([issue#17837](#), [pr#12097](#), John Spray)
- cephfs: systemd: add ceph-fuse service file ([pr#11542](#), Patrick Donnelly)
- cephfs: test fragment size limit ([issue#16164](#), [pr#1069](#), Patrick Donnelly)
- cephfs: test readahead is working ([issue#16024](#), [pr#1046](#), Patrick Donnelly)
- cephfs: test: temporarily remove fork()ing flock tests ([issue#16556](#), [pr#11211](#), John Spray)
- cephfs: tool/cephfs: displaying “list” in journal event mode ([pr#11236](#), huanwen ren)
- cephfs: tools/cephfs: add pg\_files command ([issue#17249](#), [pr#11026](#), John Spray)
- cephfs: tools/cephfs: add scan\_links command which fixes linkages errors ([pr#11446](#), Yan, Zheng)
- cephfs: update tests to enable multimds when needed ([pr#933](#), Greg Farnum)
- cleanup: build: The Light Clangtastic ([pr#11921](#), Adam C. Emerson)
- cleanup,common: common/blkdev: use realpath instead of readlink to resolve the recurs... ([pr#12462](#), Xinze Chi)
- cleanup,common: common/throttle: simplify Throttle::\_wait() ([pr#11165](#), xie xingguo)
- cleanup,common: src/common: remove nonused config option ([pr#12311](#), Wei Jin)
- cleanup: coverity fix: fixing few coverity issue ([pr#9624](#), Gaurav Kumar Garg)
- cleanup: deprecate readdir\_r() with readdir() ([pr#11805](#), Kefu Chai)
- cleanup: erasure-code: fix gf-complete warning ([pr#12150](#), Kefu Chai)

- cleanup: fix typos ([pr#12502](#), xianxiaxiao)
- cleanup: mds/FSMMap.cc: prevent unneeded copy of map entry ([pr#11798](#), Michal Jarzabek)
- cleanup: mds/FSMMap.h: add const and reference ([pr#11802](#), Michal Jarzabek)
- cleanup: mds/FSMMap: pass shared\_ptr by const ref ([pr#11383](#), Michal Jarzabek)
- cleanup: mds/SnapServer: add const to member function ([pr#11688](#), Michal Jarzabek)
- cleanup: mon/MonCap.h: add std::move for std::string ([pr#10722](#), Michal Jarzabek)
- cleanup: mon/OSDMonitor: only show interesting flags in health warning ([issue#18175](#), [pr#12365](#), Sage Weil)
- cleanup: msg/async: assert(0) -> ceph\_abort() ([pr#12339](#), Li Wang)
- cleanup: msg/AsyncMessenger: remove unneeded include ([pr#9846](#), Michal Jarzabek)
- cleanup: msg/async/rdma: fix disconnect log line ([pr#12254](#), Adir Lev)
- cleanup: msg/async: remove unused member variable ([pr#12387](#), Kefu Chai)
- cleanup: msg: fix format specifier for unsigned value id ([pr#11145](#), Weibing Zhang)
- cleanup: msg/Pipe: move DelayedDelivery class to cc file ([pr#10447](#), Michal Jarzabek)
- cleanup: msg/test: fix the guided compile-command to ceph\_test\_msgr ([pr#10490](#), Yan Jun)
- cleanup: osd/PGBackend: build\_push\_op segment fault ([pr#9357](#), Zengran Zhang)
- cleanup: osd/PG.h: change PGRecoveryStats struct to class ([pr#11178](#), Michal Jarzabek)
- cleanup: osd/PG.h: remove unneeded forward declaration ([pr#12135](#), Li Wang)
- cleanup: osd/ReplicatedPG: remove unneeded use of count ([pr#11251](#), Michal Jarzabek)
- cleanup: os/filestore: clean filestore perfcounters ([pr#11524](#), Wei Jin)
- cleanup: os/fs/FS.cc: condition on WITH\_AIO for FreeBSD ([pr#11913](#), Willem Jan Withagen)
- cleanup, rbd: cls\_rbd: silence compiler warnings ([pr#11363](#), xiexingguo)
- cleanup, rbd: journal: avoid logging an error when a watch is blacklisted ([issue#18243](#), [pr#12473](#), Jason Dillaman)

- cleanup, rbd: journal: prevent repetitive error messages after being blacklisted ([issue#18243](#), [pr#12497](#), Jason Dillaman)
- cleanup, rbd: librbd/ImageCtx: no need for virtual dtor ([pr#12220](#), Sage Weil)
- cleanup, rbd: rbd-mirror: configuration overrides for hard coded timers ([pr#11840](#), Dongsheng Yang)
- cleanup, rbd: rbd-mirror: set SEQUENTIAL and NOCACHE advise flags on image sync ([issue#17127](#), [pr#12280](#), Mykola Golub)
- cleanup: remove unneeded forward declaration ([pr#12257](#), Li Wang, Yunchuan Wen)
- cleanup: remove unused declaration ([pr#12466](#), Li Wang, Yunchuan Wen)
- cleanup, rgw: rgw multisite: move lease up to RunBucketSync instead of child crs ([pr#11598](#), Casey Bodley)
- cleanup, rgw: rgw/rest: don't print empty x-amz-request-id ([pr#10674](#), Marcus Watts)
- cleanup, rgw: verified: f23 ([pr#12103](#), Radoslaw Zarzynski)
- cleanup: src/common/perf\_counters.h: fix wrong word ([pr#11690](#), zhang.zezhu)
- cleanup: Wip ctypos ([pr#12495](#), xianxiakiao)
- cleanup: xio: provide dout\_prefix for XioConnection ([pr#9444](#), Avner BenHanoch)
- cleanup: yasm-wrapper: translate “-isystem \$1” to “-i \$1” ([pr#12093](#), Kefu Chai)
- cmake: add -Wno-unknown-pragmas to CMAKE\_CXX\_FLAGS ([pr#12128](#), Kefu Chai)
- cmake: check WITH\_RADOSGW for fcgi and expat dependencies ([pr#11481](#), David Disseldorf)
- cmake: compile C code with c99 ([pr#12369](#), Kefu Chai)
- cmake: detect keyutils if WITH\_LIBCEPHFS OR WITH\_RBD ([pr#12359](#), Kefu Chai)
- cmake: do not link erasure tests again libosd ([pr#11738](#), Kefu Chai)
- cmake: find gperftools package for tcmalloc\_minimal too ([pr#11403](#), Bassam Tabbara)
- cmake: fix boost build on ubuntu 16.10 yakkety ([pr#12143](#), Bassam Tabbara)
- cmake: Fix for cross compiling ([pr#11404](#), Bassam Tabbara)
- cmake: fix git version string, cleanup ([pr#11661](#), Sage Weil)
- cmake: librbd cleanup ([pr#11842](#), Kefu Chai)

- cmake: link tests against static librados ([issue#17260](#), [pr#11575](#), Kefu Chai)
- cmake: pass CMAKE\_BUILD\_TYPE down to rocksdb ([pr#11767](#), Kefu Chai)
- cmake: remove include/Makefile.am ([pr#11666](#), Kefu Chai)
- cmake: replace civetweb symlink w/file copy ([pr#11900](#), Matt Benjamin)
- cmake: should link against \${ALLOC\_LIBS} ([pr#11978](#), Kefu Chai)
- cmake: src/test/CMakeLists.txt: Exclude test on HAVE\_BLKID ([pr#12301](#), Willem Jan Withagen)
- cmake: Support for embedding Ceph Daemons ([pr#11764](#), Bassam Tabbara)
- cmake: use external project for rocksdb ([pr#11385](#), Bassam Tabbara)
- common: Add throttle\_get\_started perf counter ([pr#12163](#), Bartłomiej Święcki)
- common: assert(0) -> ceph\_abort() ([pr#12031](#), Sage Weil)
- common: auth: fix NULL pointer access when trying to delete CryptoAESKeyHandler instance ([pr#11614](#), runsisi)
- common,bluestore: compressor: fixes and tests; disable zlib isal (it's broken) ([pr#11349](#), Sage Weil)
- common,bluestore: mempool: mempool infrastructure, bluestore changes to use it ([pr#11331](#), Allen Samuels, Sage Weil)
- common: buffer: add advance(unsigned) back ([issue#17809](#), [pr#11993](#), Kefu Chai)
- common: buffer: add copy(unsigned, ptr) back ([issue#17809](#), [pr#12246](#), Kefu Chai)
- common: client/Client.cc: fix/silence "logically dead code" CID-Error ([pr#291](#), Yehuda Sadeh)
- common: common/strtol.cc: Get error testing also to work on FreeBSD ([pr#12034](#), Willem Jan Withagen)
- common: fix clang compilation error ([pr#12565](#), Mykola Golub)
- common: FreeBSD/EventKqueue.{h,cc} Added code to restore events on (thread)fork ([pr#11430](#), Willem Jan Withagen)
- common: log/LogClient: fill seq & who for syslog and graylog ([issue#16609](#), [pr#10196](#), Xiaoxi Chen)
- common: make l\_finisher\_complete\_lat more accurate ([pr#11637](#), Pan Liu)
- common: msg/simple/Acceptor.cc: replace shutdown() with selfpipe event in poll() (FreeBSD) ([pr#10720](#), Willem Jan Withagen)

- common: osdc/Objecter: fix relock race ([issue#17942](#), [pr#12234](#), Sage Weil)
- common: osdc/Objecter: handle race between calc\_target and handle\_osd\_map ([issue#17942](#), [pr#12055](#), Sage Weil)
- common: osd/osdmap: fix divide by zero error ([pr#12521](#), Yunchuan Wen)
- common: release g\_ceph\_context before returns ([issue#17762](#), [pr#11733](#), Kefu Chai)
- common: Remove the runtime dependency on lsb\_release ([issue#17425](#), [pr#11365](#), Brad Hubbard)
- common: test/fio: fix global CephContext life cycle ([pr#12245](#), Igor Fedotov)
- core: auth: tolerate missing MGR keys during upgrade ([pr#11401](#), Sage Weil)
- core,bluestore: os/bluestore: fix warning and uninit variable ([pr#12032](#), Sage Weil)
- core,bluestore: os: fix offsets for move\_ranges operation ([pr#11595](#), Sage Weil)
- core,bluestore: os: remove move\_ranges\_destroy\_src ([pr#11791](#), Sage Weil)
- core: ceph-disk: allow using a regular file as a journal ([issue#17662](#), [pr#11619](#), Jayashree Candadai, Loic Dachary)
- core: ceph-disk: resolve race conditions ([issue#17889](#), [issue#17813](#), [pr#12136](#), Loic Dachary)
- core,cephfs: osdc/ObjectCacher: wake up dirty stat waiters after removing buffers ([issue#17275](#), [pr#11593](#), Yan, Zheng)
- core: ceph.in: allow 'flags' to not be present in cmddescs ([issue#18297](#), [pr#12540](#), Dan Mick)
- core,cleanup: ceph-disk: do not create bluestore wal/db partitions by default ([issue#18291](#), [pr#12531](#), Loic Dachary)
- core,cleanup,common: common/TrackedOp: remove unused 'now' in \_dump() ([pr#12007](#), John Spray)
- core,cleanup: FileStore: Only verify split when it has been really done and done correctly ([pr#11731](#), Li Wang)
- core,cleanup: kv: remove snapshot iterator ([pr#12049](#), Sage Weil)
- core,cleanup: mon/MonClient.h: remove repeated searching of map ([pr#10601](#), Michal Jarzabek)
- core,cleanup: msg: Fix typos in socket creation error message ([pr#11907](#), Brad Hubbard)

- core,cleanup: osd/command tell: check pgid at the right time ([pr#11547](#), Javeme)
- core,cleanup: osd/OSDMap.cc: fix duplicated assignment for new\_blacklist\_entries ([pr#11799](#), Ker Liu)
- core,cleanup: osd/PG.cc: prevent repeated searching of map/set ([pr#11203](#), Michal Jarzabek)
- core,cleanup: osd/ReplicatedPG: remove redundant check for balance/localize read ([pr#10209](#), runsisi)
- core,cleanup: osd/ReplicatedPG: remove unneeded use of count ([pr#11242](#), Michal Jarzabek)
- core,cleanup: os/filestore: handle EINTR returned by io\_getevents() ([pr#11890](#), Pan Liu)
- core,cleanup: os/ObjectStore: remove legacy tbl support ([pr#11770](#), Jianpeng Ma)
- core,cleanup: scan build fixes ([pr#12148](#), Kefu Chai)
- core,cleanup: src: rename ReplicatedPG to PrimaryLogPG ([pr#12487](#), Samuel Just)
- core,cleanup: Wip scrub misc ([pr#11397](#), David Zafman)
- core,common: buffer: put buffers in buffer\_{data,meta} mempools ([pr#11839](#), Sage Weil)
- core,common: msg: add entity\_addr\_t types; add new entity\_addrvec\_t type ([pr#9825](#), Zhao Junwang, Sage Weil)
- core,common: msg/simple/Pipe: handle addr decode error ([issue#18072](#), [pr#12221](#), Sage Weil)
- core: compress: Fix compilation failure from missing header ([pr#12108](#), Adam C. Emerson)
- core: denc: don't pass null instances into encoder fns ([issue#17636](#), [pr#11577](#), John Spray)
- core: erasure-code: synchronize with upstream gf-complete ([issue#18092](#), [pr#12382](#), Loic Dachary)
- core: FreeBSD/OSD.cc: add client\_messenger to the avoid\_ports set. ([pr#12463](#), Willem Jan Withagen)
- core: include/object: pass "snapid\_t&" to bound\_encode() ([pr#11552](#), Kefu Chai)
- core: kv/RocksDBStore: Don't update rocksdb perf\_context if rocksdb\_perf di... ([pr#12064](#), Jianpeng Ma)
- core: librados-dev: install inline\_memory.h ([issue#17654](#), [pr#11730](#), Josh Durgin)

- core: messages/MForward: reencode forwarded message if target has differing features ([pr#11610](#), Sage Weil)
- core,mgr: messages: fix out of range assertion ([pr#11345](#), John Spray)
- core: mon,ceph-disk: add lockbox permissions to bootstrap-osd ([issue#17849](#), [pr#11996](#), Loic Dachary)
- core: mon: make it more clearly to debug for paxos state ([pr#12438](#), song baisen)
- core: mon/OSDMonitor: encode full osdmmaps with features all OSDs can understand ([pr#11284](#), Sage Weil)
- core: mon/OSDMonitor: encode OSDMap::Incremental with same features as OSDMap ([pr#11596](#), Sage Weil)
- core: mon/OSDMonitor: newly created osd should not be wrongly marked in ([pr#11795](#), runsisi)
- core: mon/OSDMonitor: remove duplicate jewel/kraken flag warning ([pr#11775](#), Josh Durgin)
- core: mon/PGMap: PGs can be stuck more than one thing ([issue#17515](#), [pr#11339](#), Sage Weil)
- core: mon: print the num\_pools and num\_objects in 'ceph -s -f json/json-p...' ([issue#17703](#), [pr#11654](#), huangjun)
- core: msg/async/AsyncConnection: dispatch write handler on keepalive2 ([issue#17664](#), [pr#11601](#), Ilya Dryomov)
- core: msg/async: DPDKStack as AsyncMessenger backend ([pr#10748](#), Haomai Wang)
- core: msg/async/rdma: change log level: 0 -> 1 ([pr#12334](#), Avner BenHanoch)
- core: msg/async/rdma: don't use more buffers than what device capabilities ... ([pr#12263](#), Avner BenHanoch)
- core: msg/async/rdma: ensure CephContext existed ([pr#12068](#), Haomai Wang)
- core: msg/async/rdma: event polling thread can block on event ([pr#12270](#), Haomai Wang)
- core: msg/async/rdma: fixup memory free ([pr#12236](#), gongchuang)
- core: msg/async/rdma: set correct value to memory manager ([pr#12299](#), Adir Lev)
- core: msg/async: set nonce before starting the workers ([pr#12390](#), Kefu Chai)
- core: msg: make loopback Connection feature accurate all the time ([pr#11183](#), Sage Weil)

- core: msg: seed random engine used for ms\_type="random" ([pr#11880](#), Casey Bodley)
- core: msg/simple/Pipe: avoid returning 0 on poll timeout ([issue#18184](#), [pr#12375](#), Sage Weil)
- core: msg/simple/Pipe::stop\_and\_wait: unlock pipe\_lock for stop\_fast\_dispatching() ([issue#18042](#), [pr#12307](#), Samuel Just)
- core: msg/simple: save the errno in case being changed by subsequent codes ([pr#10297](#), Yan Jun)
- core: osd/ECTransaction: only write out the hinfo if not delete ([issue#17983](#), [pr#12141](#), Samuel Just)
- core: OSDMonitor: only reject MOSDBoot based on up\_from if inst matches ([issue#17899](#), [pr#12003](#), Samuel Just)
- core: osd,mon: require sortbitwise flag to upgrade beyond jewel ([pr#11772](#), Sage Weil)
- core: osd/osd\_types: fix the osd\_stat\_t::decode() ([pr#12235](#), Kefu Chai)
- core: osd/PG: add "down" pg state (distinct from down+peering) ([pr#12289](#), Sage Weil)
- core: osd/PGLog::proc\_replica\_log,merge\_log: fix bound for last\_update ([issue#18127](#), [pr#12340](#), Samuel Just)
- core: osd/ReplicatedPG: do\_update\_log\_missing: take the pg lock in the callback ([issue#17789](#), [pr#11754](#), Samuel Just)
- core: osd/ReplicatedPG::record\_write\_error: don't leak orig\_reply on cancel ([issue#18180](#), [pr#12450](#), Samuel Just)
- core: os/filestore: avoid to get the wrong hardlink number. ([pr#11841](#), huangjun)
- core: os/filestore/chain\_xattr.h:uses ENODATA, so include compat.h ([pr#12279](#), Willem Jan Withagen)
- core: os/filestore: Fix erroneous WARNING: max attr too small ([issue#17420](#), [pr#11246](#), Brad Hubbard)
- core: os/FileStore: fix fiemap issue in xfs when #extents > 1364 ([pr#11554](#), Ning Yao)
- core: os/filestore: fix journal logger ([pr#12099](#), Wei Jin)
- core: os/filestore: fix potential result code overwriting ([pr#11491](#), xie xingguo)
- core: os/filestore/HashIndex: fix list\_by\_hash\_\* termination on reaching end ([issue#17859](#), [pr#11898](#), Sage Weil)

- core: os/ObjectStore: properly clear object map when replaying OP\_REMOVE ([issue#17177](#), [pr#11388](#), Yan, Zheng)
- core,performance: msg/async: ibverbs/rdma support ([pr#11531](#), Haomai Wang, Zhi Wang)
- core,performance: osd/OSDMap.cc: remove unneeded use of count ([pr#11221](#), Michal Jarzabek)
- core,performance: osd/PrimaryLogPG: don't truncate if we don't have to for WRITEFULL ([pr#12534](#), Samuel Just)
- core,performance: os/fs/FS: optimize aio::pwritev which make caller provide length. ([pr#9062](#), Jianpeng Ma)
- core,pybind,common: python-rados: implement new aio\_execute ([pr#12140](#), Iain Buclaw)
- core,rbd,bluestore,rgw,performance,cephfs: fast denc encoding ([pr#11027](#), Sage Weil)
- core: remove spurious executable permissions on source code files ([pr#1061](#), Samuel Just)
- core: ReplicatedPG::failed\_push: release read lock on failure ([issue#17857](#), [pr#11914](#), Kefu Chai)
- core: rocksdb: update to latest, and make it the default for the mons ([pr#11354](#), Sage Weil)
- core: set dumpable flag after setuid ([issue#17650](#), [pr#11582](#), Patrick Donnelly)
- core: systemd/ceph-disk: reduce ceph-disk flock contention ([issue#18049](#), [issue#13160](#), [pr#12200](#), David Disseldorp)
- core: tchaikov ([issue#17713](#), [pr#11382](#), Haomai Wang)
- core,tests: ceph\_test\_rados\_api\_tier: dump hitset that we fail to decode ([issue#17945](#), [pr#12057](#), Sage Weil)
- core,tests: common osd: Improve scrub analysis, list-inconsistent-obj output and osd-scrub-repair test ([issue#18114](#), [pr#9613](#), Kefu Chai, David Zafman)
- core,tests: test,cmake: turn unit.h into unit.cc to speed up compilation ([pr#12194](#), Kefu Chai)
- core,tests: test/rados/list.cc: Memory leak in ceph\_test\_rados\_api\_list ([issue#18250](#), [pr#12479](#), Brad Hubbard)
- core,tests: workunits/ceph-helpers.sh: Fixes for FreeBSD ([pr#12085](#), Willem Jan Withagen)

- core, tools: Added append functionality to rados tool. ([pr#11036](#), Tomy Cheru)
- core, tools: Tested-by: Huawei Ren <[ren.huanwen@zte.com.cn](mailto:ren.huanwen@zte.com.cn)> ([issue#17400](#), [pr#11276](#), Kefu Chai)
- core, tools: vstart: decrease pool size if <3 OSDs ([pr#11528](#), John Spray)
- crush: make counting of choose\_tries consistent ([issue#17229](#), [pr#10993](#), Vicente Cheng)
- crush: remove the crush\_lock ([pr#11830](#), Adam C. Emerson)
- crush: Silence coverity warnings for test/crush/crush.cc ([pr#12436](#), Brad Hubbard)
- doc: Add doc about osd scrub {during recovery|chunk {min|max}| sleep} ([pr#12176](#), Paweł Sadowski)
- doc: Add docs about looking up Monitors through DNS ([issue#14527](#), [pr#10852](#), Wido den Hollander)
- doc: add docs for raw compression ([pr#12244](#), Casey Bodley)
- doc: Add documentation about mon\_allow\_pool\_delete before pool remove ([pr#11943](#), Wido den Hollander)
- doc: add infernalis EOL date ([pr#11925](#), Ken Dreyer)
- doc: adding changelog for v10.2.4 ([pr#12346](#), Abhishek Lekshmanan)
- doc: Add MON docs about pool flags and pool removal config settings ([pr#10853](#), Wido den Hollander)
- doc: add python-rgw doc ([pr#11859](#), Kefu Chai)
- doc: change the osd\_max\_backfills default to 1 ([issue#17701](#), [pr#11658](#), huangjun)
- doc: clarify file deletion from OSD restricted pool behaviour ([issue#17937](#), [pr#12054](#), David Disseldorp)
- doc: clarify mds deactivate purpose ([pr#11957](#), Patrick Donnelly)
- doc: common/Throttle: fix typo for BackoffThrottle ([pr#12129](#), Wei Jin)
- doc: correcting the object name ([pr#12354](#), Uday Mullangi)
- doc: Correcting the sample python tempurl generation script. ([issue#15258](#), [pr#8712](#), Diwakar Goel)
- doc: Coverity and SCA fixes ([pr#7784](#), Danny Al-Gaaf)
- doc: doc/dev/osd\_internals: add pgpool.rst ([pr#12500](#), Brad Hubbard)
- doc: doc/dev/perf: a few notes on perf ([pr#12168](#), Sage Weil)

- doc: doc/dev/perf: fix dittotherapy ([pr#12317](#), xie xingguo)
- doc: doc/man: avoid file builtin to solve build error ([pr#11984](#), Patrick Donnelly)
- doc: doc/rados/configuration/ms-ref.rst: document a few async msgr options ([pr#12126](#), Piotr Dałek)
- doc: doc/rados/configuration/osd-config-ref.rst: document the fast mark down ([pr#12124](#), Piotr Dałek)
- doc: doc/release-notes: kraken release notes (draft) ([pr#12338](#), Sage Weil)
- doc: doc/releases: add links to kraken and v10.2.4 ([pr#12409](#), Kefu Chai)
- doc: doc/start/hardware-recommendations: cosmetic ([pr#10585](#), Zhao Junwang)
- doc: Documentation syntax cleanup ([pr#11784](#), John Spray)
- doc: document osd tell bench ([issue#5431](#), [pr#16](#), Sage Weil)
- doc: drop -journal-check from ceph-mds man page ([issue#17747](#), [pr#11912](#), Nathan Cutler)
- doc: explain rgw\_fcgi\_socket\_backlog in rgw/config-ref.rst ([pr#12548](#), liuchang0812)
- doc: final additions to 11.1.0-rc release notes ([pr#12448](#), Abhishek Lekshmanan)
- doc: Fix broken link for caps ([issue#17587](#), [pr#11546](#), Uday Mullangi)
- doc: fix broken links ([issue#17587](#), [pr#11518](#), Uday Mullangi)
- doc: fix dead link "Hardware Recommendations" ([pr#11379](#), Leo Zhang)
- doc: fix dead link of "os-recommendations" in troubleshooting-osd ([pr#11454](#), Leo Zhang)
- doc: Fixed mapping error in legacy mds command ([pr#11668](#), Malte Fiala)
- doc: Fix for worker arguments to cephfs-data-scan tool ([pr#12360](#), Wido den Hollander)
- doc: fix grammar/spelling in RGW sections ([pr#12329](#), Ken Dreyer)
- doc: Fixing the broken hyperlinks by pointing to correct documentation. ([pr#11617](#), Uday Mullangi)
- doc: fix librados example programs ([pr#11302](#), Alexey Sheplyakov)
- doc: fix mgr literal block rST syntax ([pr#11652](#), Ken Dreyer)
- doc: fix start development cluster operation in index.rst ([pr#11233](#), Leo Zhang)

- doc: fix the script for rebuild monitor db ([pr#11962](#), Kefu Chai)
- doc: fix typos ([pr#8751](#), Li Peng)
- doc: Flag deprecated mds commands and omit deprecated mon commands in help output ([pr#11434](#), Patrick Donnelly)
- doc: mailmap: change personal info ([pr#12310](#), Wei Jin)
- doc: mailmap updates sept ([pr#10955](#), Yann Dupont)
- doc: mds: fixup “mds bal mode” Description ([pr#12127](#), huanwen ren)
- doc: mention corresponding libvirt section in nova.conf ([pr#12584](#), Marc Koderer)
- doc: Modify documentation for mon\_osd\_down\_out\_interval ([pr#12408](#), Brad Hubbard)
- doc: network-protocol typos ([pr#9837](#), Zhao Junwang)
- doc: openstack glance mitaka uses show\_multiple\_locations ([pr#12020](#), Sébastien Han)
- doc: README.FreeBSD: update to match the bimonthly FreeBSD status report ([pr#11442](#), Willem Jan Withagen)
- doc: README: hint at where to look to diagnose test failures ([pr#11903](#), Dan Mick)
- doc: reformat SubmittingPatches with more rst syntax ([pr#11570](#), Kefu Chai)
- doc: release notes for 10.2.4 ([pr#12053](#), Abhishek Lekshmanan)
- doc: release notes for 10.2.5 ([issue#18207](#), [pr#12410](#), Loic Dachary)
- doc: release notes for 11.0.2 ([pr#11369](#), Abhishek Lekshmanan)
- doc: Remove duplicate command for Ubuntu ([pr#12186](#), chrone)
- doc: reviewed-by: John Wilkins <[jowilkin@redhat.com](mailto:jowilkin@redhat.com)> ([issue#17526](#), [pr#11352](#), Loic Dachary)
- doc: reviewed-by: John Wilkins <[jowilkin@redhat.com](mailto:jowilkin@redhat.com)> ([issue#17665](#), [pr#11602](#), Jason Dillaman)
- doc: rgw: fix a typo in S3 java api example ([pr#11762](#), Weibing Zhang)
- doc: rm “type=rpm-md” from yum repositories ([pr#10248](#), Ken Dreyer)
- doc: Small styling fix to mirror documentation ([pr#9714](#), Wido den Hollander)
- doc: src/doc: fix class names in exports.txt ([pr#12000](#), John Spray)
- doc: standardize EPEL instructions ([pr#11653](#), Ken Dreyer)

- doc: update cinder key permissions for mitaka ([pr#12211](#), Sébastien Han)
- doc: Update crush-map.rst, fix a typo mistake ([pr#11785](#), whu\_liuchang)
- doc: Update filestore xattr config documentation. ([pr#11826](#), Bartłomiej Święcki)
- doc: Update install-ceph-gateway.rst ([pr#11432](#), Hans van den Bogert)
- doc: Update keystone doc about v3 options ([pr#11392](#), Proskurin Kirill)
- doc: Update layout.rst, move commands to CODE block ([pr#11987](#), liuchang0812)
- doc: we can now run multiple MDS, so qualify warning ([issue#18040](#), [pr#12184](#), Nathan Cutler)
- fs: add snapshot tests to mds thrashing ([pr#1073](#), Yan, Zheng)
- fs: enable ceph-fuse permission checking for all pjd suites ([pr#1187](#), Greg Farnum)
- fs: fix two frag\_enable fragments ([issue#6143](#), [pr#656](#), Sage Weil)
- fs: fix up dd testing again ([issue#10861](#), [pr#373](#), Greg Farnum)
- fs: fuse\_default\_permissions = 0 for kernel build test ([pr#1109](#), Patrick Donnelly)
- fs: Mantle: A Programmable Metadata Load Balancer ([pr#10887](#), Michael Sevilla)
- fs: unify common parts of sub-suites ([issue#1737](#), [pr#1282](#), Patrick Donnelly)
- librados: Add rados\_aio\_exec to the C API ([pr#11709](#), Iain Buclaw)
- librados: add timeout to watch/notify ([pr#11378](#), Ryne Li)
- librados: do not request osd ack if no completed completion is set ([pr#11204](#), Sage Weil)
- librados: For C-API, expose LIBRADOS\_OPERATION\_FULL\_FORCE flag ([pr#9172](#), Jianpeng Ma)
- librados: improvements async IO in librados and libradosstriper ([pr#10049](#), Sébastien Ponce)
- librados: Memory leaks in object\_list\_begin and object\_list\_end ([issue#18252](#), [pr#12482](#), Brad Hubbard)
- librados: postpone cct deletion ([pr#11659](#), Kefu Chai)
- librados: remove new setxattr overload to avoid breaking the C++ ABI ([issue#18058](#), [pr#12206](#), Josh Durgin)
- librados: remove unused bufferlist from rados\_write\_op\_rmxattr ([pr#12030](#), Piotr

Dalek)

- librbd: add support for snapshot namespaces ([pr#11160](#), Victor Denisov)
- librbd: API changes to support separate data pool ([pr#11353](#), Jason Dillaman)
- librbd: batch object map updates during trim ([issue#17356](#), [pr#11510](#), Venky Shankar)
- librbd: bug fixes for optional data pool support ([pr#11960](#), Venky Shankar)
- librbd: cannot access non-primary image when mirroring force disabled ([issue#16740](#), [issue#17588](#), [pr#11568](#), Jason Dillaman)
- librbd: cls\_rbd updates for separate data pool ([issue#17422](#), [pr#11327](#), Jason Dillaman)
- librbd: default features should be negotiated with the OSD ([issue#17010](#), [pr#11808](#), Mykola Golub)
- librbd: diffs to clone's first snapshot should include parent diffs ([issue#18068](#), [pr#12218](#), Jason Dillaman)
- librbd: do not create empty object map object on image creation ([issue#17752](#), [pr#11704](#), Jason Dillaman)
- librbd: enabling/disabling rbd feature should report missing dependency ([issue#16985](#), [pr#12238](#), Gaurav Kumar Garg)
- librbd: ensure consistency groups will gracefully fail on older OSDs ([pr#11623](#), Jason Dillaman)
- librbd: exclusive lock incorrectly initialized when switching to head revision ([issue#17618](#), [pr#11559](#), Jason Dillaman)
- librbd: fix rollback if failed to disable mirroring for image ([pr#11260](#), runsisi)
- librbd: ignore error when object map is already locked by current client ([issue#16179](#), [pr#12484](#), runsisi)
- librbd: ignore notify errors on missing image header ([issue#17549](#), [pr#11395](#), Jason Dillaman)
- librbd: keep rbd\_default\_features setting as bitmask ([issue#18247](#), [pr#12486](#), Jason Dillaman)
- librbd: mark request as finished after failed refresh ([issue#17973](#), [pr#12160](#), Venky Shankar)
- librbd: minor cleanup ([pr#12078](#), Dongsheng Yang)
- librbd: new API method to force break a peer's exclusive lock ([issue#18429](#),

- issue#16988, issue#18327, pr#12889, Jason Dillaman)
- librbd: parse rbd\_default\_features config option as a string (pr#11175, Alyona Kiseleva, Alexey Sheplyakov)
  - librbd: possible assert failure creating image when using data pool (pr#11641, Venky Shankar)
  - librbd: proper check for get\_data\_pool compatibility (issue#17791, pr#11755, Mykola Golub)
  - librbd: properly order concurrent updates to the object map (issue#16176, pr#12420, Jason Dillaman)
  - librbd: release lock after demote (issue#17880, pr#11940, Mykola Golub)
  - librbd: remove consistency group rbd cli and API support (issue#18231, pr#12475, Jason Dillaman)
  - librbd: remove image header lock assertions (issue#18244, pr#12472, Jason Dillaman)
  - librbd: remove unused local variable (pr#12388, Yunchuan Wen)
  - librbd: silence the unused variable warning (pr#11678, Kefu Chai)
  - librbd: snap\_get\_limit compatibility check (pr#11766, Mykola Golub)
  - librbd: update internals to use optional separate data pool (pr#11356, Jason Dillaman)
  - librbd: use proper snapshot when computing diff parent overlap (issue#18200, pr#12396, Xiaoxi Chen)
  - log: optimize header file dependency (pr#9768, Xiaowei Chen)
  - mds: add debug assertion for issue #17636 (pr#11576, Yan, Zheng)
  - mds: add tests for mantle (programmable balancer) (pr#1145, Michael Sevilla)
  - mds: check if down mds is known (issue#17670, pr#11611, Patrick Donnelly)
  - mds: don't access mdsmap from log submit thread (issue#18047, pr#12208, Yan, Zheng)
  - mds: don't maintain bloom filters in standby replay (issue#16924, pr#12133, John Spray)
  - mds: enable rmxattr on pool\_namespace attrs (issue#17797, pr#11783, John Spray)
  - mds: fix dropping events in standby replay (issue#17954, pr#12077, John Spray)

- mds: fix EMetaBlob::fullbit xattr dump ([pr#11536](#), Sage Weil)
- mds: fix false “failing to respond to cache pressure” warning ([pr#11373](#), Yan, Zheng)
- mds: force client flush snap data before truncating objects ([issue#17193](#), [pr#11994](#), Yan, Zheng)
- mds: handle bad standby\_for\_fscids in fsmap ([issue#17466](#), [pr#11281](#), John Spray)
- mds: ignore ‘session evict’ when mds is replaying log ([issue#17801](#), [pr#11813](#), Yan, Zheng)
- mds: include legacy client fsid in FSMap print ([pr#11283](#), John Spray)
- mds: more deterministic timing on frag split/join ([issue#17853](#), [pr#12022](#), John Spray)
- mds: more unique\_pointer changes ([pr#11635](#), Patrick Donnelly)
- mds: properly commit new dirfrag before splitting it ([issue#17990](#), [pr#12125](#), Yan, Zheng)
- mds: release pool allocator memory after exceeding size limit ([issue#18225](#), [pr#12443](#), John Spray)
- mds: remove duplicated log in handle\_client\_readdir ([pr#11806](#), Zhi Zhang)
- mds: remove “-journal-check” help text ([issue#17747](#), [pr#11739](#), Nathan Cutler)
- mds: remove unused EFragment::OP\_ONESHOT ([pr#11887](#), John Spray)
- mds: repair backtraces during scrub ([issue#17639](#), [pr#11578](#), John Spray)
- mds: require MAY\_SET\_POOL to set pool\_ns ([issue#17798](#), [pr#11789](#), John Spray)
- mds: respawn using /proc/self/exe ([issue#17531](#), [pr#11362](#), Patrick Donnelly)
- mds: revert “mds/Mutation: remove redundant \_dump method” ([issue#17906](#), [pr#11985](#), Patrick Donnelly)
- mds: use parse\_filesystem in parse\_role to handle exceptions and reuse parsing code ([issue#17518](#), [pr#11357](#), Patrick Donnelly)
- mds: use projected path construction for access ([issue#17858](#), [pr#12063](#), Patrick Donnelly)
- mds: use unique\_ptr to simplify resource mgmt ([pr#11543](#), Patrick Donnelly)
- mgr: doc/mgr: fix mgr how long to wait to failover ([pr#11550](#), huanwen ren)
- mgr: init() return when connection daemons failed && add some err info ([pr#11424](#),

- huanwen ren)
- mgr: misc minor changes ([issue#17455](#), [pr#11386](#), xie xingguo)
  - mgr: PyModules.cc: remove duplicated if condition for fs\_map ([pr#11639](#), Weibing Zhang)
  - mgr: remove unnecessary C\_StdFunction ([pr#11883](#), John Spray)
  - mon: add missing space in warning message ([pr#11361](#), Patrick Donnelly)
  - mon: clean legacy code ([pr#9643](#), Wei Jin)
  - mon: clear duplicated logic in MDSMonitor ([pr#11209](#), Zhi Zhang)
  - mon: Do not allow pools to be deleted by default ([pr#11665](#), Wido den Hollander)
  - mon: fix “OSDs marked OUT wrongly after monitor failover” ([issue#17719](#), [pr#11664](#), Dong Wu)
  - mon: Forbidden copy and assignment function in monoprequest ([pr#9513](#), song baisen)
  - mon: have mon-specific features & rework internal monmap structures ([pr#10907](#), Joao Eduardo Luis)
  - mon: if crushtool config is empty use internal crush test ([pr#11765](#), Bassam Tabbara)
  - mon: make MDSMonitor tolerant of slow mon elections ([issue#17308](#), [pr#11167](#), John Spray)
  - mon: MonmapMonitor: return success when monitor will be removed ([issue#17725](#), [pr#11747](#), Joao Eduardo Luis)
  - mon: move case CEPH\_MSG\_POOLOP to OSDs group ([pr#11848](#), Javeme)
  - mon: osdmap’s epoch should be more than 0 ([pr#9859](#), Na Xie)
  - mon: OSDMonitor: fix the check error of pg creating ([issue#17169](#), [pr#10916](#), Desmonds)
  - mon: paxos add the timeout function when peon recovery ([pr#10359](#), song baisen)
  - mon: preserve osd weight when marking osd out, then in ([pr#11293](#), Sage Weil)
  - mon: prevent post-jewel OSDs from booting if require\_jewel\_osds is not set ([pr#11498](#), Sage Weil)
  - mon: remove ceph-create-keys from mon startup ([issue#16036](#), [pr#9345](#), Owen Synge)
  - mon: remove the redundant judgement in LogMonitor tick function ([pr#10474](#), song

- baisen)
- mon: remove utime\_t param in \_dump ([pr#12029](#), Patrick Donnelly)
  - mon: send updated monmap to its subscribers ([issue#17558](#), [pr#11456](#), Kefu Chai)
  - mon: small change on the HealthMonitor start\_epoch function ([pr#10296](#), songbaisen)
  - mon: support for building without leveldb + mon mkfs bug fix ([pr#11800](#), Bassam Tabbara)
  - osd: add a pg \_fastinfo attribute to reduce per-io metadata updates ([pr#11213](#), Sage Weil)
  - osd: Add config option to disable new scrubs during recovery ([issue#17866](#), [pr#11874](#), Wido den Hollander)
  - osd: a few fast dispatch optimizations ([pr#12052](#), Sage Weil)
  - osd: cleanup C\_CompleteSplits::finish() ([pr#12094](#), Jie Wang)
  - osd: clean up PeeringWQ::\_dequeue(), remove unnecessary variable ([pr#12117](#), Jie Wang)
  - osd: clean up process\_peering\_events ([pr#12009](#), Jie Wang)
  - osdc/Objecter: resend pg commands on interval change ([issue#18358](#), [pr#12910](#), Samuel Just)
  - osd: condition OSDMap encoding on features ([pr#12166](#), Sage Weil)
  - osd: default osd\_scrub\_during\_recovery=false ([pr#12402](#), Sage Weil)
  - osd: do not open pgs when the pg is not in pg\_map ([issue#17806](#), [pr#11803](#), Xinze Chi)
  - osd: drop stray debug message ([pr#11296](#), Sage Weil)
  - osd: EC Overwrites ([issue#17668](#), [pr#11701](#), Tomy Cheru, Samuel Just)
  - osd: enhance logging for osd network error ([pr#12458](#), liuchang0812)
  - osd: fix CEPH OSD FLAG\_RWORDERED ([pr#12603](#), Sage Weil)
  - osd: fix duplicated id of incompat feature "fastinfo" ([pr#11588](#), xie xingguo)
  - osd: fix ec scrub errors ([issue#17999](#), [pr#12306](#), Samuel Just)
  - osd: fixes to make rbd on ec work ([pr#12305](#), Samuel Just)
  - osd: Fix map gaps again (bug 15943) ([issue#15943](#), [pr#12571](#), Samuel Just)

- osd: fix memory leak from EC write workload ([issue#18093](#), [pr#12256](#), Sage Weil)
- osd: fix rados write op hang ([pr#11143](#), Yunchuan Wen)
- osd: Fix read error propagation in ECBackend ([issue#17966](#), [pr#12142](#), Samuel Just)
- osd: fix scrub boundary to not include a SnapSet ([pr#11255](#), Samuel Just)
- osd: fix signed/unsigned comparison warning ([pr#12400](#), Greg Farnum)
- osd: fix typo in PG:::clear\_primary\_state ([pr#11513](#), Brad Hubbard)
- osd: Fix typos in PG:::find\_best\_info ([pr#11515](#), Brad Hubbard)
- osd: fix typos in “struct OSDOp” comments ([pr#12350](#), Chanyoung Park)
- osd: Flush Journal on shutdown ([pr#11249](#), Wido den Hollander)
- osd: force watch PING to be write ordered ([issue#18310](#), [pr#12590](#), Samuel Just)
- osd: handle EC recovery read errors ([issue#13937](#), [pr#9304](#), David Zafman)
- osd: heartbeat peers need to be updated when a new OSD added into an existed cluster ([issue#18004](#), [pr#12069](#), Pan Liu)
- osd: Increase priority for inactive PGs backfill ([pr#12389](#), Bartłomiej Święcki)
- osd: kill PG\_STATE\_SPLITTING ([pr#11824](#), xie xingguo)
- osd: mark queued flag for op ([pr#12352](#), Yunchuan Wen)
- osd: osdc: pass a string reference type to “osdmap->lookup\_pg\_pool\_name” ([pr#12219](#), Leo Zhang)
- osd: osd/OSDMonitor: accept ‘osd pool set ...’ value as string ([pr#911](#), David Zafman)
- osd: PGLog: initialize writeout\_from in PGLog constructor ([issue#12973](#), [pr#558](#), Sage Weil)
- osd/PrimaryLogPG: don’t update digests for objects with mismatched names ([issue#18409](#), [pr#12803](#), Samuel Just)
- osd/PrimaryLogPG::failed\_push: update missing as well ([issue#18165](#), [pr#12911](#), Samuel Just)
- osd: print log when osd want to kill self ([pr#9288](#), Haomai Wang)
- osd: Remove extra call to reg\_next\_scrub() during splits ([issue#16474](#), [pr#11206](#), David Zafman)
- osd: remove redundant call of heartbeat\_check ([pr#12130](#), Pan Liu)

- osd: remove the lock heartbeat\_update\_lock, and change heartbeat\_need... ([pr#12461](#), Pan Liu)
- osd: remove the redundant clear method in consume\_map function ([pr#10553](#), song baisen)
- osd: Remove unused '\_lsb\_release\_' declarations ([pr#11364](#), Brad Hubbard)
- osd: replace hb\_out and hb\_in with a single hb\_peers ([issue#18057](#), [pr#12178](#), Pan Liu)
- osd: ReplicatedPG: don't bless C OSD SendMessageOnConn ([issue#13304](#), [pr#669](#), Jason Dillaman)
- osd: set server-side limits on omap get operations ([pr#12059](#), Sage Weil)
- osd: When deep-scrub errors present upgrade regular scrubs ([pr#12268](#), David Zafman)
- performance,bluestore: kv/MemDB: making memdb code adapt to generic maps ([pr#11436](#), Ramesh Chander)
- performance,bluestore: os/bluestore: allow default to buffered write ([pr#11301](#), Sage Weil)
- performance,bluestore: os/bluestore: bluestore\_cache\_meta\_ratio = .5 ([pr#11919](#), Sage Weil)
- performance,bluestore: os/bluestore: reduce Onode in-memory footprint ([pr#12568](#), Igor Fedotov)
- performance,bluestore: os/bluestore: refactor bluestore\_sync\_submit\_transaction ([pr#11537](#), Sage Weil)
- performance,bluestore: os/bluestore: speed up omap-key generation for same onode(the read paths) ([pr#11894](#), xie xingguo)
- performance,bluestore: os/bluestore: speedup the performance of multi-replication flow by switc... ([pr#11844](#), Pan Liu)
- performance,cephfs: Fix long stalls when calling ceph\_fsync() ([issue#17563](#), [pr#11710](#), Jeff Layton)
- performance,cleanup: Context: std::move the callback param in FunctionContext's ctor ([pr#11892](#), Kefu Chai)
- performance,cleanup: osd/PG.h: move shared ptr instead of copying it ([pr#11154](#), Michal Jarzabek)
- performance,common: common/config\_opts.h: Optimized RocksDB WAL settings. ([pr#11530](#), Mark Nelson)

- performance, common: osd/OSDMap: improve the performance of pg\_to\_acting\_osds ([pr#12190](#), Pan Liu)
- performance: msg/async: set ms\_async\_send\_inline to false to improve small randread iops ([pr#11521](#), Mark Nelson)
- performance, tools: rados: add hints to rados bench ([pr#12169](#), Sage Weil)
- pybind: avoid “exception ‘int’ object is not iterable” ([pr#11532](#), Javeme)
- pybind,cephfs: ceph\_volume\_client: fix recovery from partial auth update ([issue#17216](#), [pr#11304](#), Ramana Raja)
- pybind,cephfs: ceph\_volume\_client: set an existing auth ID’s default mon caps ([issue#17800](#), [pr#11917](#), Ramana Raja)
- pybind: ceph-rest-api: understand the new style entity\_addr\_t representation ([issue#17742](#), [pr#11686](#), Kefu Chai)
- pybind: clean up mgr stuff for flake8 ([pr#11314](#), John Spray)
- pybind: fix build failure of rgwfile binding ([pr#11825](#), Kefu Chai)
- pybind: pybind/rados: add missing “length” requires for aio\_execute() ([pr#12439](#), Kefu Chai)
- pybind: pybind/rados: Add @requires for all aio methods ([pr#12327](#), Iain Buclaw)
- qa: fixed distros links ([pr#12773](#), Yuri Weinstein)
- qa: Fixed link to centos distro ([pr#12768](#), Yuri Weinstein)
- qa/suites: switch from centos 7.2 to 7.x ([pr#12632](#), Sage Weil)
- qa/tasks/peer: update task based on current peering behavior ([issue#18330](#), [pr#12614](#), Sage Weil)
- qa/tasks/workunit: clear clone dir before retrying checkout ([issue#18336](#), [pr#12630](#), Sage Weil)
- qa: update Ubuntu image url after ceph.com refactor ([issue#18542](#), [pr#12960](#), Jason Dillaman)
- qa/workunits/rbd/test\_lock\_fence.sh fails ([issue#18388](#), [pr#12752](#), Nathan Cutler)
- rbd: added rbd-nbd fsx test case ([pr#1049](#), Jason Dillaman)
- rbd: add fsx journal replay test case ([pr#821](#), Jason Dillaman)
- rbd: add singleton to assert no rbdmap regression ([issue#14984](#), [pr#902](#), Nathan Cutler)

- rbd: add some missing workunits ([pr#870](#), Josh Durgin)
- rbd: add support for separate image data pool ([issue#17424](#), [pr#11355](#), Jason Dillaman)
- rbd: expose rbd unmap options ([issue#17554](#), [pr#11370](#), Ilya Dryomov)
- rbd: fix json formatting for image and journal status output ([issue#18261](#), [pr#12525](#), Mykola Golub)
- rbd: fix parsing of group and image specific pools ([pr#11632](#), Victor Denisov)
- rbd: journal: do not prematurely flag object recorder as closed ([issue#17590](#), [pr#11520](#), Jason Dillaman)
- rbd: krbd: kernel client expects ip[:port], not an entity\_addr\_t ([pr#11902](#), Ilya Dryomov)
- rbd: -max\_part and -nbds\_max options for nbd map ([issue#18186](#), [pr#12379](#), Pan Liu)
- rbd: move nbd test workload to separate client host from OSDs ([pr#1170](#), Jason Dillaman)
- rbd: provision volumes to format as XFS ([issue#6693](#), [pr#1028](#), Loic Dachary)
- rbd: rbd-mirror: fix sparse read optimization in image sync ([issue#18146](#), [pr#12368](#), Mykola Golub)
- rbd: rbd-mirror HA: move librbd::image\_watcher::Notifier to librbd::object\_watcher ([issue#17017](#), [pr#11290](#), Mykola Golub)
- rbd: rbd-mirror: recovering after split-brain ([issue#16991](#), [issue#18051](#), [pr#12212](#), Mykola Golub)
- rbd: rbd-mirror: snap protect of non-layered image results in split-brain ([issue#16962](#), [pr#11744](#), Mykola Golub)
- rbd: rbd-nbd: disallow mapping images >2TB in size ([issue#17219](#), [pr#11741](#), Mykola Golub)
- rbd: rbd-nbd: invalid error code for “failed to read nbd request” messages ([issue#18242](#), [pr#12483](#), Mykola Golub)
- rbd: rbd-nbd: restart parent process logger after forking ([issue#18070](#), [pr#12222](#), Jason Dillaman)
- rbd: rbd-nbd: support disabling auto-exclusive lock transition logic ([issue#17488](#), [pr#11438](#), Mykola Golub)
- rbd: rbd-nbd: support partition for rbd-nbd mapped raw block device ([issue#18115](#), [pr#12259](#), Pan Liu)

- rbd: tests with rbd\_skip\_partial\_discard option enabled ([pr#1077](#), Mykola Golub)
- rbd,tools: rbd : make option -stripe-unit w/ B/K/M work ([pr#12407](#), Jianpeng Ma)
- rbd: updated tests to use new rbd default feature set ([pr#842](#), Jason Dillaman)
- rbd: use snap\_remove implementation from internal ([pr#12035](#), Victor Denisov)
- rgw: add default zone name ([issue#7009](#), [pr#954](#), Orit Wasserman)
- rgw: add documentation for upgrading with rgw\_region\_root\_pool ([pr#12138](#), Orit Wasserman)
- rgw: add option to log custom HTTP headers (rgw\_log\_http\_headers) ([pr#7639](#), Matt Benjamin)
- rgw: add recovery procedure for upgrade to older version of jewel ([issue#17820](#), [pr#11827](#), Orit Wasserman)
- rgw: add rgw\_compression\_type=random for teuthology testing ([pr#11901](#), Casey Bodley)
- rgw: add sleep to let the sync agent init ([pr#1136](#), Orit Wasserman)
- rgw: add suport for creating S3 type subuser of admin rest api ([issue#16682](#), [pr#10325](#), snakeAngel2015)
- rgw: add support for the prefix parameter in account listing of Swift API ([issue#17931](#), [pr#12047](#), Radoslaw Zarzynski)
- rgw: allow fastcgi idle timeout to be adjusted ([pr#230](#), Sage Weil)
- rgw: also approve, passed teuthology (many false positives in several classes) ([issue#17985](#), [pr#12224](#), Yehuda Sadeh, Sage Weil)
- rgw: Anonymous users shouldn't be able to access requester pays buckets. ([issue#17175](#), [pr#11719](#), Zhang Shaowen)
- rgw: aws4: add presigned url bugfix in runtime ([issue#16463](#), [pr#10160](#), Javier M. Mellid)
- rgw: bucket resharding ([issue#17550](#), [pr#11230](#), Yehuda Sadeh)
- rgw:bugfix for deleting objects name beginning and ending with underscores of one bucket using POST method of AWS's js sdk. ([issue#17888](#), [pr#11982](#), root)
- rgw: Class member cookie is not initialized correctly in some coroutine's constructor. ([pr#11673](#), Zhang Shaowen)
- rgw: clean up RGWShardedOmapCRManager on early return ([issue#17571](#), [pr#11505](#), Casey Bodley)

- rgw: clear data\_sync\_cr if RGWDataSyncControlCR fails ([issue#17569](#), [pr#11506](#), Casey Bodley)
- rgw: compilation of the ASIO front-end is enabled by default. ([pr#12073](#), Radoslaw Zarzynski)
- rgw: compression uses optional::emplace instead of in-place factories ([pr#12021](#), Radoslaw Zarzynski)
- rgw: conform to the standard usage of string::find ([pr#10086](#), Yan Jun)
- rgw: data\_extra\_pool is unique per zone ([issue#17025](#), [pr#1119](#), Orit Wasserman)
- rgw: delete entries\_index in RGWFetchAllMetaCR ([issue#17812](#), [pr#11816](#), Casey Bodley)
- rgw: do not abort when accept a CORS request with short origin ([pr#12381](#), LiuYang)
- rgw: do not enable both tcp and uds for fastcgi ([issue#5797](#), [pr#479](#), Andrew Schoen)
- rgw: don't error out on empty owner when setting acls ([issue#6892](#), [pr#877](#), Loic Dachary, Nathan Cutler)
- rgw: Don't loop forever when reading data from 0 sized segment. ([issue#17692](#), [pr#11567](#), Marcus Watts)
- rgw: dont set CURLOPT\_UPLOAD for GET requests ([issue#17822](#), [pr#12105](#), Casey Bodley)
- rgw: don't store empty chains in gc ([issue#17897](#), [pr#11969](#), Yehuda Sadeh)
- rgw: do quota tests on ubuntu ([issue#6382](#), [pr#635](#), Sage Weil)
- rgw: dump objects in RGWBucket::check\_object\_index() ([issue#14589](#), [pr#11324](#), Yehuda Sadeh)
- rgw: dump remaining coroutines when cr deadlock is detected ([pr#11580](#), Casey Bodley)
- rgw: extract host name from host:port string ([issue#17788](#), [pr#11751](#), Yehuda Sadeh)
- rgw: Fixed problem with PUT with x-amz-copy-source when source object is compressed. ([pr#12253](#), Adam Kupczyk)
- rgw: fixes for virtual hosting of buckets ([issue#17440](#), [issue#15975](#), [issue#17136](#), [pr#11280](#), Casey Bodley, Robin H. Johnson)
- rgw: fix etag in multipart complete ([issue#17794](#), [issue#6830](#), [issue#16129](#), [issue#17872](#), [pr#1269](#), Casey Bodley, Orit Wasserman)

- rgw: fix for bucket delete racing with mdlog sync ([issue#17698](#), [pr#11648](#), Casey Bodley)
- rgw: fix for passing temporary in InitBucketSyncStatus ([issue#17661](#), [pr#11594](#), Casey Bodley)
- rgw: fix for unsafe change of rgw\_zonegroup ([issue#17962](#), [pr#12075](#), Casey Bodley)
- rgw: fix indentation for cache\_pools ([issue#8295](#), [pr#251](#), Sage Weil)
- rgw: fix missing master zone for a single zone zonegroup ([issue#17364](#), [pr#11965](#), Orit Wasserman)
- rgw: fix osd crashes when execute "radosgw-admin bi list -max-entries=1" command ([issue#17745](#), [pr#11697](#), weiqiaomiao)
- rgw: fix put\_acls for objects starting and ending with underscore ([issue#17625](#), [pr#11566](#), Orit Wasserman)
- rgw: fix RGWSimpleRadosLockCR set\_description() ([pr#11961](#), Tianshan Qu)
- rgw: fix the field 'total\_time' of log entry in log show opt ([issue#17598](#), [pr#11425](#), weiqiaomiao)
- rgw: fix uncompressed object size deduction in RGWRados::copy\_obj\_data. ([issue#17803](#), [pr#11794](#), Radoslaw Zarzynski)
- rgw: frontend subsystem rework ([pr#10767](#), Radoslaw Zarzynski, Casey Bodley, Matt Benjamin)
- rgw: ftw ([issue#17888](#), [pr#12262](#), Casey Bodley)
- rgw: get\_system\_obj does not use result of get\_system\_obj\_state ([issue#17580](#), [pr#11444](#), Casey Bodley)
- rgw: get\_zonegroup() uses "default" zonegroup if empty ([issue#17372](#), [pr#11207](#), Yehuda Sadeh)
- rgw: handle empty POST condition ([issue#17635](#), [pr#11581](#), Yehuda Sadeh)
- rgw: handle Swift auth errors in a way compatible with new Tempests. ([issue#16590](#), [pr#10021](#), Radoslaw Zarzynski)
- rgw: json encode/decode index\_type, allow modification ([issue#17755](#), [pr#11707](#), Yehuda Sadeh)
- rgw: loses realm/period/zonegroup/zone data: period overwritten if somewhere in the cluster is still running Hammer ([issue#17371](#), [pr#11426](#), Orit Wasserman)
- rgw: make RGWLocalAuthApplier::is\_admin\_of() aware about system users. ([issue#18106](#), [pr#12283](#), Radoslaw Zarzynski)

- rgw: metadata sync info should be shown at master zone of slave zone... ([issue#18091](#), [pr#12187](#), Jing Wenjun)
- rgw: minor cleanup ([pr#10057](#), Yan Jun)
- rgw: move compression config into zone placement ([pr#12113](#), Casey Bodley)
- rgw: move xfs to a seperate directory ([pr#969](#), Orit Wasserman)
- rgw: multipart upload copy ([issue#12790](#), [pr#11269](#), Yehuda Sadeh, Javier M. Mellid)
- rgw: need to close\_section in lc list op ([pr#12232](#), weiqiaomiao)
- rgw: policy acl format should be xml ([pr#946](#), Orit Wasserman)
- rgw: radosgw-admin: more on placement configuration ([issue#18078](#), [pr#12242](#), Casey Bodley)
- rgw: region conversion respects pre-existing rgw\_region\_root\_pool ([issue#17963](#), [pr#12076](#), Casey Bodley)
- rgw: remove a redundant judgement when listng objects. ([pr#10849](#), zhangshaowen)
- rgw: remove circular reference in RGWAsyncRadosRequest ([issue#17793](#), [issue#17792](#), [pr#11815](#), Casey Bodley)
- rgw: remove suggestion to upgrade libcurl ([pr#11630](#), Casey Bodley)
- rgw: remove unused variable “ostr” in rgw\_b64.h and fix the comment ([pr#11329](#), Weibing Zhang)
- rgw: Replacing ‘+’ with “%20” in canonical uri for s3 v4 auth. ([issue#17076](#), [pr#10919](#), Pritha Srivastava)
- rgw: revert unintentional change to civetweb ([pr#12004](#), Bassam Tabbara)
- rgw: rgw-admin: new commands to control placement ([issue#18078](#), [pr#12230](#), Yehuda Sadeh)
- rgw: RGWBucketSyncStatusManager uses existing async\_rados ([issue#18083](#), [pr#12229](#), Casey Bodley)
- rgw: rgw\_file: apply missed base64 try-catch ([issue#17663](#), [pr#11671](#), Matt Benjamin)
- rgw: RGWHTTPArgs::get\_str() - return argument string that was set. ([pr#10672](#), Marcus Watts)
- rgw: rgw multisite: fix the increamtal bucket sync init ([issue#17624](#), [pr#11553](#), Zengran Zhang)

- rgw: rgw multisite: use a rados lock to coordinate data log trimming ([pr#10546](#), Casey Bodley)
- rgw: RGW Python bindings - use explicit array ([pr#11831](#), Daniel Gryniewicz)
- rgw: rgw\_rados.cc fix shard\_num format for sprintf ([pr#11493](#), Weibing Zhang)
- rgw: rgw/rgw\_file.cc: Add compat.h to allow CLOCK\_MONOTONE ([pr#12309](#), Willem Jan Withagen)
- rgw: RGWSimpleRadosReadCR tolerates empty reads ([issue#17568](#), [pr#11504](#), Casey Bodley)
- rgw: [RGW] Wip rgw compression ([pr#11494](#), Alyona Kiseleva, Adam Kupczyk, Casey Bodley)
- rgw: set duration for lifecycle lease ([issue#17965](#), [pr#12231](#), Yehuda Sadeh)
- rgw: should assign 'olh\_bh' to state.attrset[RGW\_ATTR\_OLH\_ID\_TAG] instead of 'bh' ([pr#10239](#), weiqiaomiao)
- rgw: skip empty http args in method parse() to avoid extra effort ([pr#11989](#), Guo Zhandong)
- rgw: split osd's in 2 nodes ([issue#15612](#), [pr#1019](#), Vasu Kulkarni)
- rgw: support for x-robots-tag header ([issue#17790](#), [pr#11753](#), Yehuda Sadeh)
- rgw: sync modules, metadata search ([pr#10731](#), Yehuda Sadeh)
- rgw: Update version of civetweb to 1.8 ([pr#11343](#), Marcus Watts)
- rgw: use civetweb if no frontend was configured ([pr#958](#), Orit Wasserman)
- rgw: use explicit flag to cancel RGWCoroutinesManager::run() ([issue#17465](#), [pr#12452](#), Casey Bodley)
- rgw: valgrind fixes for kraken ([issue#18414](#), [issue#18407](#), [issue#18412](#), [issue#18300](#), [pr#12949](#), Casey Bodley)
- rgw: verified that failed check is in osd-scrub-repair.sh ([issue#17850](#), [pr#11881](#), Matt Benjamin)
- rgw: we don't support btrfs any more ([pr#1132](#), Orit Wasserman)
- rgw: Wip rgwfile pybind ([pr#11624](#), Haomai Wang)
- tests,bluestore: os/bluestore: add UT for an estimation of Onode in-memory size ([pr#12532](#), Igor Fedotov)
- tests,bluestore: os/test/store\_test: fix legacy bluestore cache settings application ([pr#11915](#), Igor Fedotov)

- tests: ceph-disk: force debug monc = 0 ([issue#17607](#), [pr#11534](#), Loic Dachary)
- tests: ceph\_objectstore\_tool.py: Don't use btrfs on FreeBSD ([pr#10507](#), Willem Jan Withagen)
- tests: ceph\_test\_objectstore: fix Rename test ([pr#12261](#), Sage Weil)
- tests: check hostname -fqdn sanity before running make check ([issue#18134](#), [pr#12297](#), Nathan Cutler)
- tests,cleanup,rbd: test/librbd: in test\_notify set object-map and fast-diff features by default ([pr#11821](#), Mykola Golub)
- tests,cleanup: test\_bloom\_filter.cc: Fix a mismatch for the random\_seed parameter ([pr#11774](#), Willem Jan Withagen)
- tests,cleanup: test/osd/osd-fast-mark-down.sh: remove unnecessary teardown() calls ([pr#12101](#), Kefu Chai)
- tests,cleanup: test/osd-scrub-repair.sh: use repair() instead of "ceph pg repair" ([pr#12036](#), Kefu Chai)
- tests,cleanup: test/rados: remove unused bufferlist variable ([pr#10221](#), Yan Jun)
- tests,common: test: add perf-reset test in test/perf\_counters.cc ([pr#8948](#), wangsongbo)
- tests: disable failing tests ([issue#17561](#), [issue#17757](#), [pr#11714](#), Loic Dachary)
- tests: disable the echo when running get\_timeout\_delays() ([pr#12180](#), Kefu Chai)
- tests: do not use memstore.test\_temp\_dir in two tests ([issue#17743](#), [pr#12281](#), Loic Dachary)
- tests: erasure-code: add k=2, m=2 for isa & jerasure ([issue#18188](#), [pr#12383](#), Loic Dachary)
- tests: facilitate background process debug in ceph-helpers.sh ([issue#17830](#), [pr#12183](#), Loic Dachary)
- tests: fix ceph-helpers.sh wait\_for\_clean delays ([issue#17830](#), [pr#12095](#), Loic Dachary)
- tests: fix osd-scrub-repair.sh ([pr#12072](#), David Zafman)
- tests: Fix racey test by setting noout flag (tracker 17757) ([issue#17757](#), [pr#11715](#), David Zafman)
- tests: merge ceph-qa-suite
- tests: Minor clean-ups ([pr#12048](#), David Zafman)

- tests: minor make check cleanup ([pr#12146](#), David Zafman)
- tests: no python3 tests for ceph-disk ([issue#17923](#), [pr#12025](#), Loic Dachary)
- tests: osd-crush.sh must retry crush dump ([issue#17919](#), [pr#12016](#), Loic Dachary)
- tests: osd-scrub-repair.sh abort if add\_something fails ([pr#12172](#), Loic Dachary)
- tests: os/memstore: fix a mem leak in MemStore::Collection::create\_object() ([pr#12201](#), Kefu Chai)
- tests: os/memstore, os/filestore: fix store\_test's to satisfy rm\_coll behavi... ([pr#11558](#), Igor Fedotov)
- tests: paxos fixes ([issue#11913](#), [pr#457](#), John Spray)
- tests: pin flake8 to avoid behavior changes ([issue#17898](#), [pr#11971](#), Loic Dachary)
- tests: qa: fixed script to schedule rados and other suites with -subset option ([pr#12587](#), Yuri Weinstein)
- tests: qa/tasks/admin\_socket: subst in repo name ([pr#12508](#), Sage Weil)
- tests: qa/tasks/ceph\_deploy: use dev option instead of dev-commit ([pr#12514](#), Vasu Kulkarni)
- tests: qa/tasks/osd\_backfill.py: wait for osd.[12] to start ([issue#18303](#), [pr#12577](#), Sage Weil)
- tests: qa/workunits/cephtool/test.sh: FreeBSD has no distro. ([pr#11702](#), Willem Jan Withagen)
- tests: qa/workunits: include extension for nose tests ([pr#12572](#), Sage Weil)
- tests: qa/workunits/rados/test\_envlibrados\_for\_rocksdb: force librados-dev install ([pr#11941](#), Sage Weil)
- tests,rbd: qa/workunits/rbd: fix ([issue#18271](#), [pr#12511](#), Sage Weil)
- tests,rbd: qa/workunits/rbd: removed qemu-io test case 077 ([issue#10773](#), [pr#12366](#), Jason Dillaman)
- tests,rbd: qa/workunits/rbd: simplify running nbd test under build env ([pr#11781](#), Mykola Golub)
- tests,rbd: qa/workunits/rbd: use image id when probing for image presence ([issue#18048](#), [pr#12195](#), Mykola Golub)
- tests,rbd: qa/workunits/rbd: use more recent qemu-io tests that support Xenial ([issue#18149](#), [pr#12371](#), Jason Dillaman)
- tests,rbd: rbd-mirror: fix gmock warnings in bootstrap request unit tests

- ([issue#18156](#), [pr#12344](#), Mykola Golub)
- tests,rbd: rbd-mirror: improve resiliency of stress test case ([issue#17416](#), [pr#11326](#), Jason Dillaman)
  - tests,rbd: test: new librbd discard after write test case ([pr#11645](#), Jason Dillaman)
  - tests,rbd: test: skip TestLibRBD.DiscardAfterWrite if skip partial discard enabled ([issue#17750](#), [pr#11703](#), Jason Dillaman)
  - tests,rbd: test: TestJournalReplay test cases need to wait for event commit ([issue#17566](#), [pr#11480](#), Jason Dillaman)
  - tests: remove TestPGLog EXPECT\_DEATH tests ([issue#18030](#), [pr#12361](#), Loic Dachary)
  - tests: save 9 characters for asok paths ([issue#16014](#), [pr#12066](#), Loic Dachary)
  - tests: sync ceph-erasure-code-corpus for using 'arch' not 'uname -p' ([pr#12024](#), Kefu Chai)
  - tests: test/ceph\_crypto: do not read ceph.conf in global\_init() ([issue#18128](#), [pr#12318](#), Kefu Chai)
  - tests: test: ceph-objectstore-tool: should import platform before using it ([pr#12038](#), Kefu Chai)
  - tests: test/ceph\_test\_msgr: do not use Message::middle for holding transient... ([issue#17728](#), [pr#11680](#), Kefu Chai)
  - tests: test: disable osd-scrub-repair and test-erasure-eio ([issue#17830](#), [pr#12058](#), Loic Dachary, Dan Mick)
  - tests: test: disable osd-scrub-repair and test-erasure-eio ([pr#11979](#), Dan Mick)
  - tests: test: Don't write to a poolid that this test might not have created ([pr#12378](#), David Zafman)
  - tests: test: enable unittest\_dns\_resolve ([pr#12209](#), Kefu Chai)
  - tests: test/encoding/readable.sh: fix shell script warning ([pr#11527](#), Willem Jan Withagen)
  - tests: TestErasureCodePluginJerasure must stop the log thread ([issue#17561](#), [pr#11721](#), Loic Dachary)
  - tests: test: fix test-erasure-eio and osd-scrub-repair races (17830) ([pr#11926](#), David Zafman)
  - tests: test/osd-fast-mark-down.sh: wrong assumption on first subtest ([pr#12123](#), Piotr Dałek)

- tests: test/osd/osd-fast-mark-down.sh: introduce large timeout ([issue#17918](#), [pr#12019](#), Piotr Dałek)
- tests: test/osd-scrub-repair.sh: Use test case specific object names to help... ([pr#11449](#), David Zafman)
- tests: test/store\_test: fix errors on the whole test suite run caused by the... ([pr#11427](#), Igor Fedotov)
- tests: test\_subman.sh: Don't use -tmpdir ([pr#11384](#), Willem Jan Withagen)
- tests: test: test-erasure-eio.sh fix recovery testing and enable it ([pr#12170](#), David Zafman)
- tests: The default changed to disallow pool delete as of #11665; the tests assume it's allowed. ([pr#11897](#), Sage Weil)
- tests: Turn off tests again due to Jenkins failures ([pr#12217](#), David Zafman)
- tests: unittest\_throttle avoid ASSERT\_DEATH ([issue#18036](#), [pr#12393](#), Loic Dachary)
- tests: update rbd/singleton/all/formatted-output.yaml to support ceph-ci ([issue#18440](#), [pr#12823](#), Nathan Cutler)
- tests: use shorter directories for tests ([issue#16014](#), [pr#12046](#), Loic Dachary)
- tests: vstart.sh: fix bashism in the script ([pr#11889](#), Mykola Golub)
- tests: workunits/ceph-helpers.sh: FreeBSD returns a different errorstring. ([pr#12005](#), Willem Jan Withagen)
- tools: Adding ceph-lazy tool ([pr#11055](#), gcharot)
- tools: ceph-create-keys should not try forever to do things ([issue#17753](#), [issue#12649](#), [issue#16255](#), [pr#11749](#), Alfredo Deza)
- tools: ceph\_detect\_init: add support for Alpine ([pr#8316](#), John Coyle)
- tools: ceph-disk: fix flake8 errors ([issue#17898](#), [pr#11973](#), Ken Dreyer)
- tools: ceph-disk: prevent unnecessary tracebacks from subprocess.check\_call ([issue#16125](#), [pr#12414](#), Alfredo Deza)
- tools: ceph-post-file: single command to upload a file to cephdrop ([pr#505](#), Dan Mick, Travis Rhoden)
- tools: cleanup phase of cephfs-data-scan ([pr#12337](#), Vishal Kanaujia)
- tools: osdmaptool: additional tests ([pr#1196](#), Sage Weil)
- tools: osdmaptool: fix divide by zero error ([pr#12561](#), Yunchuan Wen)

- tools: rados: fix segfaults when run without -pool ([issue#17684](#), [pr#11633](#), David Disseldorp)
- tools: rados: optionally support reading omap key from file ([issue#18123](#), [pr#12286](#), Jason Dillaman)
- tools: script/run-coverity: update ([pr#12162](#), Sage Weil)
- tools: script/sepio\_bt.sh: a script to prepare for debugging on [teuthology@sepio](#) ([pr#12012](#), Kefu Chai)
- tools: src/vstart.sh: Only execute btrfs if it is available ([pr#11683](#), Willem Jan Withagen)
- tools: tools/ceph-monstore-update-crush.sh: FreeBSD getopt is not compatible... ([pr#11525](#), Willem Jan Withagen)

## v11.0.2 Kraken

---

This development checkpoint release includes a lot of changes and improvements to Kraken. This is the first release introducing ceph-mgr, a new daemon which provides additional monitoring & interfaces to external monitoring/management systems. There are also many improvements to bluestore, RGW introduces sync modules, copy part for multipart uploads and metadata search via elastic search as a tech preview.

# Notable Changes

- bluestore: os/bluestore: misc fixes ([pr#10953](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: do not op\_file\_update deleted files ([pr#10686](#), Sage Weil)
- bluestore: bluestore/BitAllocator: Fix deadlock with musl libc ([pr#10634](#), John Coyle)
- bluestore: bluestore/BlueFS: revert direct IO for WRITER\_WAL ([pr#11059](#), Mark Nelson)
- bluestore: ceph-disk: support creating block.db and block.wal with customized size for bluestore ([pr#10135](#), Zhi Zhang)
- bluestore: compressor/zlib: switch to raw deflate ([pr#11122](#), Piotr Dałek)
- bluestore: do not use freelist to track bluefs\_extents ([pr#10698](#), Sage Weil)
- bluestore: initialize csum\_order properly ([pr#10728](#), xie xingguo)
- bluestore: kv/rocksdb: dump transactions on error ([pr#11042](#), Somnath Roy)
- bluestore: kv: In memory keyvalue db implementation ([pr#9933](#), Ramesh Chander)
- bluestore: os/bluestore/BitAllocator: batch is\_allocated bit checks ([pr#10704](#), Ramesh Chander)
- bluestore: os/bluestore/BlueFS: For logs of rocksdb & bluefs only use directio. ([pr#11012](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: async compaction ([pr#10717](#), Varada Kari, Sage Weil)
- bluestore: os/bluestore/BlueFS: do not hold internal lock while waiting for IO ([pr#9898](#), Varada Kari, Sage Weil)
- bluestore: os/bluestore/BlueFS: do not start racing async compaction ([pr#11010](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: don't inc l\_bluefs\_files\_written\_wal if overwrite. ([pr#10143](#), Jianpeng Ma)
- bluestore: os/bluestore/BlueFS: factor unflushed log into runway calculation ([pr#10966](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: fix async compaction logging bug ([pr#10964](#), Sage Weil)

- bluestore: os/bluestore/BlueFS: log dirty files at sync time ([pr#11108](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: only extend extent on same bdev ([pr#11023](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: prevent concurrent async compaction ([pr#11095](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: release completed aios ([pr#11268](#), Sage Weil)
- bluestore: os/bluestore/BlueFS: use StupidAllocator; fix async compaction bug ([pr#11087](#), Sage Weil)
- bluestore: os/bluestore/bluefs: add file refs check ([pr#10863](#), xie xingguo)
- bluestore: os/bluestore/bluefs: use map to track dirty files ([pr#10923](#), xie xingguo)
- bluestore: os/bluestore/bluefs\_types: fix extent operator<< ([pr#10685](#), Sage Weil)
- bluestore: os/bluestore/bluestore\_types: uint64\_t for ref\_map ([pr#11267](#), Sage Weil)
- bluestore: os/bluestore: Hint based allocation in bitmap Allocator ([pr#10978](#), Ramesh Chander)
- bluestore: os/bluestore: Remove bit alloc Woverloaded-virtual warnings ([pr#10082](#), Ramesh Chander)
- bluestore: os/bluestore: a few cleanups ([pr#11192](#), xie xingguo)
- bluestore: os/bluestore: a few fixes about the global csum setting ([pr#11195](#), xie xingguo)
- bluestore: os/bluestore: add assert to compress\_extent\_map ([pr#11240](#), Sage Weil)
- bluestore: os/bluestore: add cache-related stats ([pr#10961](#), xie xingguo)
- bluestore: os/bluestore: add checks and kill unreachable code ([pr#11077](#), xie xingguo)
- bluestore: os/bluestore: add error injection ([pr#11151](#), Sage Weil)
- bluestore: os/bluestore: add max blob size; fix compressed min blob size logic ([pr#11239](#), Sage Weil)
- bluestore: os/bluestore: add multiple finishers to bluestore ([pr#10780](#), Ilsoo Byun)
- bluestore: os/bluestore: add perf counters for compression effectiveness and space utilization measurements ([pr#10449](#), Igor Fedotov)

- bluestore: os/bluestore: apply “small encoding” for onode\_t::extents map ([pr#10018](#), Igor Fedotov)
- bluestore: os/bluestore: avoid blob\_t reencode when unchanged ([pr#10768](#), Sage Weil)
- bluestore: os/bluestore: binary search specified shard ([pr#11245](#), xie xingguo)
- bluestore: os/bluestore: change algorithm of compression header from string to int ([pr#10137](#), xie xingguo)
- bluestore: os/bluestore: compaction fixes ([pr#11279](#), Sage Weil)
- bluestore: os/bluestore: drop redundant call of get\_blob ([pr#11275](#), xie xingguo)
- bluestore: os/bluestore: drop unreferenced spanning blobs ([pr#11212](#), Sage Weil)
- bluestore: os/bluestore: fix a few leaks ([pr#11068](#), Sage Weil)
- bluestore: os/bluestore: fix a few memory utilization leaks and wasters ([pr#11011](#), Sage Weil)
- bluestore: os/bluestore: fix crash in decode\_some() ([pr#11312](#), Sage Weil)
- bluestore: os/bluestore: fix decoding hash of bnode ([pr#10773](#), xie xingguo)
- bluestore: os/bluestore: fix fsck() won’t catch stray shard sometimes ([pr#11219](#), xie xingguo)
- bluestore: os/bluestore: fix gc when blob extends past eof ([pr#11282](#), Sage Weil)
- bluestore: os/bluestore: fix improper local var variable in collection\_list meth... ([pr#10680](#), Igor Fedotov)
- bluestore: os/bluestore: fix incorrect pool decoding of bnode ([pr#10117](#), xie xingguo)
- bluestore: os/bluestore: fix leak of result-checking of \_fsck\_check\_extents ([pr#11040](#), xie xingguo)
- bluestore: os/bluestore: fix leaks in our use of rocksdb ([pr#11250](#), Sage Weil)
- bluestore: os/bluestore: fix memory leak during bit\_alloc testing ([pr#9935](#), xie xingguo)
- bluestore: os/bluestore: fix offset bug in \_do\_write\_small. ([pr#11030](#), amoxic)
- bluestore: os/bluestore: fix onode cache addition race ([pr#11300](#), Sage Weil)
- bluestore: os/bluestore: fix potential access violation ([pr#10362](#), xie xingguo)
- bluestore: os/bluestore: fix potential access violation during rename ([pr#11033](#),

- xie xingguo)
- bluestore: os/bluestore: fix shard\_info::dump() ([pr#11061](#), xie xingguo)
- bluestore: os/bluestore: fix spanning blob leak from ~ExtentMap ([pr#11223](#), Somnath Roy)
- bluestore: os/bluestore: fix statfs tests ([pr#10910](#), Sage Weil)
- bluestore: os/bluestore: fix when block device is not a multiple of the block size ([pr#10844](#), Sage Weil)
- bluestore: os/bluestore: fix write\_big counter and some more cleanups ([pr#11344](#), xie xingguo)
- bluestore: os/bluestore: fix/improve csum error message ([pr#10938](#), Sage Weil)
- bluestore: os/bluestore: garbage collect partially overlapped blobs ([pr#11232](#), Roushan Ali)
- bluestore: os/bluestore: get rid off "isa-l" type in ZLibCompressor ctor ([pr#10931](#), xie xingguo)
- bluestore: os/bluestore: gifting bluefs more carefully ([pr#10950](#), xie xingguo)
- bluestore: os/bluestore: honour allow-eio flag; use global compressor if possible ([pr#10970](#), xie xingguo)
- bluestore: os/bluestore: improve required compression threshold ([pr#10080](#), xie xingguo)
- bluestore: os/bluestore: include bluefs space in statfs result ([pr#10795](#), Sage Weil)
- bluestore: os/bluestore: introduce power 2 macros for block alignment and rounding ([pr#10128](#), xie xingguo)
- bluestore: os/bluestore: make assert conditional with macro for allocator ([pr#11014](#), Ramesh Chander)
- bluestore: os/bluestore: make cache settings process-wide ([pr#11295](#), Sage Weil)
- bluestore: os/bluestore: make clone\_range copy-on-write ([pr#11106](#), Sage Weil)
- bluestore: os/bluestore: make onode keys more efficient (and sort correctly) ([pr#11009](#), xie xingguo, Sage Weil)
- bluestore: os/bluestore: make trim() of 2Q cache more fine-grained ([pr#9946](#), xie xingguo)
- bluestore: os/bluestore: make zone/span size of bitmap-allocator configurable ([pr#10040](#), xie xingguo)

- bluestore: os/bluestore: misc cleanup and test fixes ([pr#11346](#), Igor Fedotov)
- bluestore: os/bluestore: misc cleanups ([pr#10201](#), xie xingguo)
- bluestore: os/bluestore: misc cleanups ([pr#11197](#), Haomai Wang)
- bluestore: os/bluestore: misc fixes ([pr#9999](#), xie xingguo)
- bluestore: os/bluestore: misc fixes ([pr#10771](#), xie xingguo)
- bluestore: os/bluestore: misc. fixes ([pr#11129](#), xie xingguo)
- bluestore: os/bluestore: more cleanups ([pr#11235](#), xie xingguo)
- bluestore: os/bluestore: more cleanups and fixes ([pr#11210](#), xie xingguo)
- bluestore: os/bluestore: narrow condition of sanity check when get\_object\_key() ([pr#11149](#), xie xingguo)
- bluestore: os/bluestore: narrow lock scope for cache trim() ([pr#10410](#), xie xingguo)
- bluestore: os/bluestore: optimize intrusive sets for size. ([pr#11319](#), Mark Nelson)
- bluestore: os/bluestore: pack a few more in-memory types ([pr#11328](#), Sage Weil)
- bluestore: os/bluestore: precondition rocksdb/bluefs during mkfs ([pr#10814](#), Sage Weil)
- bluestore: os/bluestore: prevent extent merging across shard boundaries ([pr#11216](#), Sage Weil)
- bluestore: os/bluestore: print bluefs\_extents in hex ([pr#10689](#), Sage Weil)
- bluestore: os/bluestore: proper handling for csum enable/disable settings ([pr#10431](#), Igor Fedotov)
- bluestore: os/bluestore: refactor dirty blob tracking along with some related fixes ([pr#10215](#), Igor Fedotov)
- bluestore: os/bluestore: remove cmake warning from extent alloc functions ([issue#16766](#), [pr#10492](#), Ramesh Chander)
- bluestore: os/bluestore: remove deferred\_csum machinery ([pr#11243](#), Sage Weil)
- bluestore: os/bluestore: remove some copy-pastes ([pr#11017](#), Igor Fedotov)
- bluestore: os/bluestore: replace store with logger in Cache ([pr#10969](#), xie xingguo)
- bluestore: os/bluestore: shard extent map ([pr#10963](#), Sage Weil)

- bluestore: os/bluestore: simplify LRU Cache::trim() ([pr#10109](#), xie xingguo)
- bluestore: os/bluestore: simplify calculation of collection key range ([pr#11166](#), xie xingguo)
- bluestore: os/bluestore: sloppy reshards boundaries to avoid spanning blobs ([pr#11263](#), Sage Weil)
- bluestore: os/bluestore: still more cleanups ([pr#11274](#), xie xingguo)
- bluestore: os/bluestore: switch spanning\_blob\_map to std::map ([pr#11336](#), Sage Weil)
- bluestore: os/bluestore: trim cache on reads ([pr#10095](#), Sage Weil)
- bluestore: os/bluestore: try to split blobs instead of spanning them ([pr#11264](#), Sage Weil)
- bluestore: os/bluestore: upgrade compression settings to atomics ([pr#11244](#), xie xingguo)
- bluestore: os/bluestore: use small encoding for bluefs extent and fnode ([pr#10375](#), xie xingguo)
- bluestore: os/bluestore: yet another statfs test fix ([pr#10926](#), Igor Fedotov)
- bluestore: os/bluestore: Fix size calculation in bitallocator ([pr#10377](#), Ramesh Chander)
- bluestore: os/bluestore: fix error handling of posix\_fallocate() ([pr#10277](#), xie xingguo)
- bluestore: os/bluestore: use BE for gifting and reclaiming from bluefs ([pr#10294](#), xie xingguo)
- bluestore: os/bluestore: get rid off blob's ref\_map for non-shared objects ([pr#9988](#), Igor Fedotov)
- bluestore: kv/MemDB: fix wrong output target and add sanity checks ([pr#10358](#), xie xingguo)
- bluestore: os/bluestore: add a boundary check of cache read ([pr#10349](#), xie xingguo)
- bluestore: os/bluestore: fix bitmap allocating failure if max\_alloc\_size is 0 ([pr#10379](#), xie xingguo)
- bluestore: os/bluestore: misc fixes ([pr#10327](#), xie xingguo)
- bluestore: kv/MemDB: misc fixes and cleanups ([pr#10295](#), xie xingguo)
- bluestore: rocksdb: pull up to master (4.12 + a few patches) ([pr#11069](#), Sage Weil)

Weil)

- bluestore: test/store\_test: extend Bluestore compression test to verify compress... ([pr#11080](#), Igor Fedotov)
- bluestore: test/store\_test: fix statfs results check to consider SSD min\_alloc\_size ([pr#11096](#), Igor Fedotov)
- bluestore: unittest\_bluestore\_types: a few more types for sizeof ([pr#11323](#), Sage Weil)
- bluestore: ceph\_test\_objectstore: test clone\_range and fix a few bugs ([pr#11103](#), Sage Weil)
- bluestore: kv: fix some bugs in memdb ([pr#10550](#), Haodong Tang)
- bluestore: os/bluestore/BlueFS: disable buffered io ([pr#10766](#), Sage Weil)
- build/ops,bluestore: test/objectstore/CMakeLists.txt: fix libaio conditional ([pr#11008](#), Sage Weil)
- build/ops,cephfs: client: added def for ACCESSPERMS when undefined ([pr#9835](#), John Coyle)
- build/ops,cephfs: deb: merge ceph-fs-common into ceph-common ([issue#16808](#), [pr#10433](#), Nathan Cutler)
- build/ops,cephfs: man/Makefile-client.am: drop legacy cephfs tool ([pr#10444](#), Nathan Cutler)
- build/ops,cephfs: test: break out librados-using cephfs test ([issue#16556](#), [pr#10452](#), John Spray)
- build/ops,common: common/dns\_resolve: use ns\_name\_uncompress instead of ns\_name\_ntop ([pr#9755](#), John Coyle)
- build/ops,common: msg/async/net\_handler.cc: make it more compatible with BSDs ([pr#10029](#), Willem Jan Withagen)
- build/ops,pybind: Include Python 3 bindings into the cmake build and make packages for them ([pr#10208](#), Oleh Prypin)
- build/ops,rbd: systemd: add install section to rbdmap.service file ([pr#10942](#), Jelle vd Kooij)
- build/ops,rbd: test: fix rbd-mirror workunit test cases for cmake ([pr#10076](#), Jason Dillaman)
- build/ops,rgw: rgw-ldap: add ldap lib to rgw lib deps based on build config ([pr#9852](#), John Coyle)
- build/ops: .gitignore: Add .pyc files globally ([pr#11076](#), Brad Hubbard)

- build/ops: Allow compressor build without YASM ([pr#10937](#), Daniel Gryniewicz)
- build/ops: CMake - stop pip checking for updates ([pr#10161](#), Daniel Gryniewicz)
- build/ops: CMakeList.txt: link ceph\_objectstore\_tool against fuse only if WITH\_FUSE ([pr#10149](#), Willem Jan Withagen)
- build/ops: Cmake: fix using CMAKE\_DL\_LIBS instead of dl ([pr#10317](#), Willem Jan Withagen)
- build/ops: CmakeLists.txt: use LIB\_RESOLV instead of resolv. ([pr#10972](#), Willem Jan Withagen)
- build/ops: Enable builds without ceph-test subpackage ([issue#16776](#), [pr#10872](#), Ricardo Dias)
- build/ops: Fix libatomic\_ops-devel in SUSE and specfile cleanup ([issue#16645](#), [pr#10363](#), Nathan Cutler)
- build/ops: FreeBSD: Define CLOCK\_REALTIME\_COARSE in compat.h ([pr#10506](#), Willem Jan Withagen)
- build/ops: Gentoo support for ceph-disk / ceph-detect-init; pip speedup ([pr#8317](#), Robin H. Johnson)
- build/ops: LTTng-UST disabled for openSUSE ([issue#16937](#), [pr#10592](#), Michel Normand)
- build/ops: Port ceph-brag to Python 3 (+ small fixes) ([pr#10064](#), Oleh Prypin)
- build/ops: Removes remaining reference to WITH\_MDS ([pr#10286](#), J. Eric Ivancich)
- build/ops: Stop hiding errors from run-tox.sh ([issue#17267](#), [pr#11071](#), Dan Mick)
- build/ops: Wip kill warnings ([pr#10881](#), Kefu Chai)
- build/ops: autogen: Fix rocksdb error when make dist ([pr#10988](#), tianqing)
- build/ops: autotools: remove a few other remaining traces ([pr#11019](#), Sage Weil)
- build/ops: build scripts: Enable dnf for Fedora >= 22 ([pr#11105](#), Brad Hubbard)
- build/ops: build: drop dryrun of autogen.sh from run-cmake-check.sh script ([pr#11013](#), xie xingguo)
- build/ops: ceph-disk tests: Let missing python interpreters be non-fatal ([pr#11072](#), Dan Mick)
- build/ops: ceph-disk: Compatibility fixes for Python 3 ([pr#9936](#), Anirudha Bose)
- build/ops: ceph-disk: do not activate device that is not ready ([issue#15990](#), [pr#9943](#), Boris Ranto)

- build/ops: ceph-osd-prestart.sh: check existence of OSD data directory ([issue#17091](#), [pr#10809](#), Nathan Cutler)
- build/ops: ceph-osd-prestart.sh: drop Upstart-specific code ([issue#15984](#), [pr#9667](#), Nathan Cutler)
- build/ops: ceph-post-file replace DSA with RSA ssh key ([issue#14267](#), [pr#10800](#), David Galloway)
- build/ops: ceph.spec.in: don't try to package \_\_pycache\_\_ for SUSE ([issue#17106](#), [pr#10805](#), Tim Serong)
- build/ops: ceph.spec.in: fix rpm package building error ([pr#10115](#), runsisi)
- build/ops: changes for Clang and yasm ([pr#10417](#), Willem Jan Withagen)
- build/ops: cmake changes ([pr#10351](#), Kefu Chai)
- build/ops: cmake changes ([pr#10059](#), Kefu Chai)
- build/ops: cmake changes ([pr#10279](#), Kefu Chai)
- build/ops: cmake changes ([issue#16804](#), [pr#10391](#), Kefu Chai)
- build/ops: cmake changes ([pr#10361](#), Kefu Chai)
- build/ops: cmake changes ([pr#10112](#), Kefu Chai)
- build/ops: cmake changes ([pr#10489](#), Kefu Chai)
- build/ops: cmake changes ([pr#10283](#), Kefu Chai)
- build/ops: cmake changes ([issue#16504](#), [pr#9995](#), Kefu Chai, Sage Weil, Dan Mick)
- build/ops: cmake changes ([pr#9975](#), Kefu Chai)
- build/ops: cmake changes related to LTTng-UST ([pr#10917](#), Kefu Chai)
- build/ops: common/compressor: add libcommon as a dependency for zlib and snappy p... ([pr#11083](#), Igor Fedotov)
- build/ops: compat: add abstractions for non portable pthread name funcs ([pr#9763](#), John Coyle)
- build/ops: configure.ac: Use uname instead of arch. ([pr#9766](#), John Coyle)
- build/ops: configure.ac: add \_LIBS variables for boost\_system and boost\_iostreams ([pr#9848](#), John Coyle)
- build/ops: configure.ac: fix res\_query detection ([pr#9820](#), John Coyle)
- build/ops: debian and cmake cleanups ([pr#10788](#), Kefu Chai)

- build/ops: debian: bump compat to 9 ([issue#16744](#), [pr#10366](#), Kefu Chai)
- build/ops: debian: python related changes ([pr#10322](#), Kefu Chai)
- build/ops: debian: replace SysV rbdmap with systemd service ([pr#10435](#), Ken Dreyer)
- build/ops: debian: set libexec dir to correct value as autotools did ([pr#10096](#), Daniel Grynewicz)
- build/ops: do\_cmake.sh: set up initial plugin dir ([pr#10067](#), Sage Weil)
- build/ops: fix /etc/os-release parsing in install-deps.sh ([pr#10981](#), Nathan Cutler)
- build/ops: fix the rpm build for centos ([pr#10289](#), Oleh Prypin, Josh Durgin)
- build/ops: force Python 3 packages to build in SUSE ([issue#17106](#), [pr#10894](#), Dominique Leuenberger, Nathan Cutler)
- build/ops: install-deps.sh based on /etc/os-release ([issue#16522](#), [pr#10017](#), Jan Fajerski)
- build/ops: install-deps: exit non-zero when we cannot match distro ([pr#10941](#), Gregory Meno)
- build/ops: isa-l: add isa-l library as a submodule ([pr#10066](#), Alyona Kiseleva)
- build/ops: jerasure: include generic objects in neon jerasure lib (like sse3/4) ([pr#10879](#), Dan Mick)
- build/ops: logrotate: Run as root/ceph ([pr#10587](#), Boris Ranto)
- build/ops: lttng: build the tracepoint provider lib from .c files in repo ([pr#11196](#), Kefu Chai)
- build/ops: make-dist: generate ceph.spec ([issue#16501](#), [pr#9986](#), Sage Weil)
- build/ops: make-dist: set rpm\_release correctly for release builds ([pr#11334](#), Dan Mick)
- build/ops: make-srpm.sh: A simple script to make the srpm for ceph. ([pr#11064](#), Ira Cooper)
- build/ops: makefile: change librgw\_file\_\* as check\_PROGRAMS ([issue#16646](#), [pr#10229](#), Brad Hubbard)
- build/ops: remove autotools ([pr#11007](#), Sage Weil)
- build/ops: rpm: Do not start targets on update ([pr#9968](#), Nathan Cutler, Boris Ranto)

- build/ops: rpm: ExclusiveArch for suse\_version ([issue#16936](#), [pr#10594](#), Michel Normand)
- build/ops: rpm: Fix creation of mount.ceph symbolic link for SUSE distros ([pr#10353](#), Ricardo Dias)
- build/ops: rpm: add udev BuildRequires to provide /usr/lib/udev directory ([issue#16949](#), [pr#10608](#), Nathan Cutler)
- build/ops: rpm: build rpm with cmake ([pr#10016](#), Kefu Chai)
- build/ops: rpm: drop obsolete libs-compat and python-ceph-compat metapackages ([issue#16353](#), [pr#9757](#), Nathan Cutler)
- build/ops: rpm: fix permissions for /etc/ceph/rbdmap ([issue#17395](#), [pr#11217](#), Ken Dreyer)
- build/ops: rpm: fix shared library devel package names and dependencies ([issue#16345](#), [issue#16346](#), [pr#9744](#), Nathan Cutler, Ken Dreyer)
- build/ops: rpm: move mount.ceph from ceph-base to ceph-common and add symlink in /sbin for SUSE ([issue#16598](#), [pr#10147](#), Nathan Cutler)
- build/ops: run-cmake-check.sh: Remove redundant calls ([pr#11116](#), Brad Hubbard)
- build/ops: script: improve ceph-release-notes regex ([pr#10729](#), Nathan Cutler)
- build/ops: src/CMakeLists.txt: remove double flag -Wno-invalid-offsetof ([pr#10443](#), Willem Jan Withagen)
- build/ops: src/CMakeLists.txt: remove unneeded libraries from ceph-dencoder target ([pr#10478](#), Willem Jan Withagen)
- build/ops: src/global/pidfile.cc: Assign elements in structures individually ([pr#10516](#), Willem Jan Withagen)
- build/ops: src/kv/CMakeLists.txt: force rocksdb/include to first include directory ([pr#11194](#), Willem Jan Withagen)
- build/ops: test/common/test\_util.cc: FreeBSD does not have distro information ([pr#10547](#), Willem Jan Withagen)
- build/ops: test: make check using cmake ([pr#10116](#), Kefu Chai, Sage Weil)
- build/ops: verified f23 ([pr#10222](#), Kefu Chai)
- build/ops: yasm-wrapper: dont echo the yasm command line ([pr#10819](#), Casey Bodley)
- build/ops: .gitignore: exclude core dumps, logfiles and temporary testresults ([pr#8150](#), Willem Jan Withagen)
- build/ops: this fixes the broken build ([pr#9992](#), Haomai Wang)

- build/ops: mrgw: search for cmake build dir. ([pr#10180](#), Abhishek Lekshmanan)
- build/ops: mrun, mstart.sh, mstop.sh: search for cmake build directory ([pr#10097](#), Yehuda Sadeh)
- build/ops: arm64 fixes([pr#10438](#), Dan Mick)
- build/ops: Wip kill warnings ([pr#10934](#), Kefu Chai)
- build/ops: systemd: add osd id to service description ([pr#10091](#), Ruben Kerkhof)
- build/ops: fix wrong indent caused compile warning ([pr#10014](#), Wanlong Gao)
- build/ops: ceph-detect-init: fix the py3 test ([pr#10266](#), Kefu Chai)
- build/ops: ceph.spec: fix ceph-mgr version requirement ([pr#11285](#), Sage Weil)
- build/ops: make-dist/ceph.spec.in: Fix srpm build breakage. ([pr#10404](#), Ira Cooper)
- build/ops: master: remove SYSTEMD\_RUN from initscript ([issue#16440](#), [issue#7627](#), [pr#9871](#), Vladislav Odintsov)
- build/ops: rocksdb: revert the change introduced by dc41731 ([pr#10595](#), Kefu Chai)
- build/ops: do\_freebsd\*.sh: rename do\_freebsd-cmake.sh to do\_freebsd.sh ([pr#11088](#), Kefu Chai)
- build/ops: gcc 6.1.1 complains about missing include: <random>. 4.8.3 does not c... ([pr#10747](#), Daniel Oliveira)
- build/ops: selinux: Allow ceph to manage tmp files ([issue#17436](#), [pr#11259](#), Boris Ranto)
- build/ops: selinux: allow read /proc/<pid>/cmdline ([issue#16675](#), [pr#10339](#), Kefu Chai)
- cephfs,common: osdc/Journaler: move C\_DelayFlush class to .cc ([pr#10744](#), Michal Jarzabek)
- cephfs,core,rbd: ObjectCacher: fix bh\_read\_finish offset logic ([issue#16002](#), [pr#9606](#), Greg Farnum)
- cephfs,core,rbd: osdc/ObjectCacher: move C\_ReadFinish, C\_RetryRead ([pr#10781](#), Michal Jarzabek)
- cephfs: Add ceph\_ll\_setlk and ceph\_ll\_getlk ([pr#9566](#), Frank S. Filz)
- cephfs: CephFS: misc. cleanups and remove legacy cephfs tool ([issue#16195](#), [issue#16035](#), [issue#15923](#), [pr#10243](#), John Spray)
- cephfs: Clean up handling of “..” in ceph client ([pr#10691](#), Jeff Layton)

- cephfs: Client: fixup param type and return value ([pr#10463](#), gongchuang)
- cephfs: Client: pass "UserPerm" struct everywhere for security checks ([issue#16367](#), [issue#17368](#), [pr#11218](#), Greg Farnum)
- cephfs: First pile of statx patches ([pr#10922](#), Sage Weil, Jeff Layton)
- cephfs: Fix attribute handling at lookup time ([issue#16668](#), [pr#10386](#), Jeff Layton)
- cephfs: Inotable repair during forward scrub ([pr#10281](#), Vishal Kanaujia)
- cephfs: Server: drop locks and auth pins if wait for pending truncate ([pr#9716](#), xie xingguo)
- cephfs: Small interface cleanups for struct ceph\_statx ([pr#11093](#), Jeff Layton)
- cephfs: build ceph-fuse on OSX ([pr#9371](#), Yan, Zheng)
- cephfs: ceph-fuse: link to libtcmalloc or jemalloc ([issue#16655](#), [pr#10258](#), Yan, Zheng)
- cephfs: ceph\_volume\_client: store authentication metadata ([issue#15406](#), [issue#15615](#), [pr#9864](#), John Spray, Ramana Raja)
- cephfs: client/barrier: move C\_Block\_Sync class to .cc ([pr#11001](#), Michal Jarzabek)
- cephfs: client/filer: cleanup the redundant judgments of \_write&&\_fallocate ([pr#10062](#), huanwen ren)
- cephfs: client: add missing client\_lock for get\_root ([pr#10027](#), Patrick Donnelly)
- cephfs: client: discard mds map if it is identical to ours ([pr#9774](#), xie xingguo)
- cephfs: client: fast abort if underlying statsf() call failed; end scope of std::hex properly ([pr#9803](#), xie xingguo)
- cephfs: client: fix access violation ([pr#9793](#), xie xingguo)
- cephfs: client: fix readdir vs fragmentation race ([issue#17286](#), [pr#11147](#), Yan, Zheng)
- cephfs: client: fix segment fault in Client::\_invalidate\_kernel\_dcache(). ([issue#17253](#), [pr#11170](#), Yan, Zheng)
- cephfs: client: fix shutdown with open inodes ([issue#16764](#), [pr#10419](#), John Spray)
- cephfs: client: include COMPLETE and ORDERED states in cache dump ([pr#10485](#), Greg Farnum)
- cephfs: client: kill compiling warning ([pr#9994](#), xie xingguo)

- cephfs: client: misc fixes ([pr#9838](#), xie xingguo)
- cephfs: client: move Inode specific cleanup to destructor ([pr#10168](#), Patrick Donnelly)
- cephfs: client: note order of member init in cons ([pr#10169](#), Patrick Donnelly)
- cephfs: client: properly set inode number of created inode in replay request ([issue#17172](#), [pr#10957](#), Yan, Zheng)
- cephfs: client: protect InodeRef with client\_lock ([issue#17392](#), [pr#11225](#), Yan, Zheng)
- cephfs: doc/mds: fixup mds doc ([pr#10573](#), huanwen ren)
- cephfs: fuse\_ll: fix incorrect error settings of fuse\_ll\_mkdir() ([pr#9809](#), xie xingguo)
- cephfs: include/ceph\_fs.h: guard #define CEPH\_SETATTR\_\* with #ifndef ([pr#10265](#), Kefu Chai)
- cephfs: libcephfs: Fix the incorrect integer conversion in libcephfs\_jni.cc ([pr#10640](#), wenjunhuang)
- cephfs: libcephfs: add umount function in cephfs.pyx ([pr#10774](#), huanwen ren)
- cephfs: libcephfs: fix portability-related error settings ([pr#9794](#), xie xingguo)
- cephfs: libcephfs: kill compiling warning ([pr#10622](#), xie xingguo)
- cephfs: mds/CDir: remove the part of judgment for \_next\_dentry\_on\_set ([pr#10476](#), zhang.zezhu)
- cephfs: mds/CInode: fix potential fin hanging ([pr#9773](#), xie xingguo)
- cephfs: mds/MDBalancer: cleanup ([pr#10512](#), huanwen ren)
- cephfs: mds/MDCache: kill a compiler warning ([pr#11254](#), xie xingguo)
- cephfs: mds/MDSMap default metadata pool to -1 (was: output None instead of 0 when no fs present.) ([issue#16588](#), [pr#10202](#), Xiaoxi Chen)
- cephfs: mds/MDSTable: add const to member functions ([pr#10846](#), Michal Jarzabek)
- cephfs: mds/SessionMap.h: change statement to assertion ([pr#11289](#), Michal Jarzabek)
- cephfs: mds/SnapRealm.h: add const to member functions ([pr#10878](#), Michal Jarzabek)
- cephfs: mds/server: clean up handle\_client\_open() ([pr#11120](#), huanwen ren)

- cephfs: mon/MDSMonitor: move C\_Updated class to .cc file ([pr#10668](#), Michal Jarzabek)
- cephfs: osdc/mds: fixup pos parameter in the journaler ([pr#10200](#), huanwen ren)
- cephfs: reduce unnecessary mds log flush ([pr#10393](#), Yan, Zheng)
- cephfs: tools/cephfs: Remove cephfs-data-scan tmap\_upgrade ([issue#16144](#), [pr#10100](#), Douglas Fuller)
- cephfs: ceph\_fuse: use sizeof get the buf length ([pr#11176](#), LeoZhang)
- cli: retry when the mon is not configured ([issue#16477](#), [pr#11089](#), Loic Dachary)
- cmake: Add -pie to CMAKE\_EXE\_LINKER\_FLAGS ([pr#10755](#), Tim Serong)
- cmake: Fix FCGI include directory ([pr#9983](#), Tim Serong)
- cmake: Fix mismatched librgw VERSION / SOVERSION ([pr#10754](#), Tim Serong)
- cmake: FreeBSD specific excludes in CMakeLists.txt ([pr#10973](#), Willem Jan Withagen)
- cmake: FreeBSD specific excludes in CMakeLists.txt files ([pr#10517](#), Willem Jan Withagen)
- cmake: Really add FCGI\_INCLUDE\_DIR to include\_directories for rgw ([pr#10139](#), Tim Serong)
- cmake: Removed README.cmake.md, edited README.md ([pr#10028](#), Ali Maredia)
- cmake: Support tcmalloc\_minimal allocator ([pr#11111](#), Bassam Tabbara)
- cmake: add dependency from ceph\_smalllobenchrd to cls libraries ([pr#10870](#), J. Eric Ivancich)
- cmake: add\_subdirectory(include) ([pr#10360](#), Kefu Chai)
- cmake: ceph\_test\_rbd\_mirror does not require librados\_test\_stub ([pr#10164](#), Jason Dillaman)
- cmake: cleanup Findgperftools.cmake ([pr#10670](#), Kefu Chai)
- cmake: correct ceph\_test\_librbd/ceph\_test\_rbd\_mirror linkage ([issue#16882](#), [pr#10598](#), Jason Dillaman)
- cmake: disable -fvar-tracking-assignments for ceph\_dencoder.cc ([pr#10275](#), Kefu Chai)
- cmake: disable unittest\_async\_compressor ([pr#10394](#), Kefu Chai)
- cmake: do not link against unused objects or libraries ([pr#10837](#), Kefu Chai)

- cmake: enable ccache for rocksdb too ([pr#11100](#), Bassam Tabbara)
- cmake: exclude non-public symbols in shared libraries ([issue#16556](#), [pr#10472](#), Kefu Chai)
- cmake: fix incorrect dependencies to librados ([pr#10145](#), Jason Dillaman)
- cmake: fix the FTBFS introduced by dc8b3ba ([pr#10282](#), Kefu Chai)
- cmake: fix the build of unittest\_async\_compressor ([pr#10400](#), Kefu Chai)
- cmake: fix the tracing header dependencies ([pr#10906](#), Kefu Chai)
- cmake: fix unittest\_rbd\_mirror failures under non-optimized builds ([pr#9990](#), Jason Dillaman)
- cmake: fix wrong path introduced by bb163e9 ([pr#10643](#), Kefu Chai)
- cmake: fixes ([pr#10092](#), Daniel Gryniewicz)
- cmake: fixes for pypi changes ([pr#10204](#), Kefu Chai)
- cmake: include(SIMDExt) in src/CMakeLists.txt ([pr#11003](#), Kefu Chai)
- cmake: install ceph\_test\_cls\_rgw ([pr#10025](#), Kefu Chai)
- cmake: install ceph\_test\_rados\_striper\_api\_\* ([pr#10541](#), Kefu Chai)
- cmake: install platlib into a subdir of build-base dir ([pr#10666](#), Kefu Chai)
- cmake: make py3 a nice-to-have ([issue#17103](#), [pr#11015](#), Kefu Chai)
- cmake: pass -DINTEL\* to gf-complete cflags ([pr#10956](#), tone.zhang, Kefu Chai)
- cmake: pass cmake's compiler and flags to compile RocksDB into build ([pr#10418](#), Willem Jan Withagen)
- cmake: recompile erasure src for different variants ([pr#10772](#), Kefu Chai)
- cmake: remove WITH\_MDS option ([pr#10186](#), Ali Maredia)
- cmake: remove more autotools hacks ([pr#11229](#), Sage Weil)
- cmake: remove unnecessary linked libs from libcephfs ([issue#16556](#), [pr#10081](#), Kefu Chai)
- cmake: rework NSS and SSL ([pr#9831](#), Matt Benjamin)
- cmake: set ARM\_CRC\_FLAGS from the CRC test rather than ARM\_NEON\_FLAGS ([issue#17250](#), [pr#11028](#), Dan Mick)
- cmake: specify distutils build path explicitly ([pr#10568](#), Kefu Chai)

- cmake: supress more warnings ([pr#10469](#), Willem Jan Withagen)
- cmake: use PERF\_LOCAL\_FLAGS only if defined ([issue#17104](#), [pr#10828](#), Michel Normand)
- cmake: use stock Find\* modules. ([pr#10178](#), Kefu Chai)
- cmake: work to get initial FreeBSD stuff ([pr#10352](#), Willem Jan Withagen)
- cmake: find GIT\_VER variables if there is no .git dir ([pr#11499](#), Ali Maredia)
- common,bluestore: Isa-l extention for zlib compression plugin ([pr#10158](#), Alyona Kiseleva, Dan Mick)
- common,bluestore: compressor/zlib: zlib wrapper fix ([pr#11079](#), Igor Fedotov)
- common: auth/cephx: misc fixes ([pr#9679](#), xie xingguo)
- common: common/PluginRegistry: improve error output for shared library load fa... ([pr#11081](#), Igor Fedotov)
- common: common/Throttle.h: remove unneeded class ([pr#10902](#), Michal Jarzabek)
- common: common/Timer.h: delete copy constr and assign op ([pr#11046](#), Michal Jarzabek)
- common: common/WorkQueue: add std move ([pr#9729](#), Michal Jarzabek)
- common: compressor: zlib compressor plugin cleanup ([pr#9782](#), Alyona Kiseleva)
- common: erasure-code: Runtime detection of SIMD for jerasure and shec ([pr#11086](#), Bassam Tabbara)
- common: global: log which process/command sent a signal ([pr#8964](#), song baisen)
- common: include/assert: clean up ceph assertion macros ([pr#9969](#), Sage Weil)
- common: instantiate strict\_si\_cast<long> not strict\_si\_cast<int64\_t> ([issue#16398](#), [pr#9934](#), Kefu Chai)
- common: lockdep: verbose even if no logging is set ([pr#10576](#), Willem Jan Withagen)
- common: messages/MOSDMap: mark as enlighten OSDMap encoder ([pr#10843](#), Sage Weil)
- common: mon/Monitor.cc:replice lock/unlock with Mutex:Lockr ([pr#9792](#), Michal Jarzabek)
- common: msg/AsyncMessenger.cc: remove code duplication ([pr#10030](#), Michal Jarzabek)
- common: msg/async: less verbose debug messages at debug\_ms=1 ([pr#11205](#), Sage

Weil)

- common: msg/async: remove static member variable ([issue#16686](#), [pr#10440](#), Kefu Chai)
- common: only call crypto::init once per CephContext ([issue#17205](#), [pr#10965](#), Casey Bodley)
- common: osdc/ObjectCacher: change iterator to const\_iterator and add const to member functions ([pr#9644](#), Michal Jarzabek)
- common: preforker: prevent call to 'write' on an fd that was already closed ([pr#10949](#), Avner BenHanoch)
- common: remove basename() dependency ([pr#9845](#), John Coyle)
- common: src/common/buffer.cc fix judgment for lseek ([pr#10130](#), zhang.zezhu)
- common: unknown hash type of judgment modification ([pr#9510](#), huanwen ren)
- common: Timer.cc: replace long types with auto ([pr#11067](#), Michal Jarzabek)
- common: TrackedOp: move ShardedTrackingData to .cc ([pr#10639](#), Michal Jarzabek)
- common: config\_opts: fix comment(radio -> ratio) ([pr#10783](#), xie xingguo)
- common: src/common/dns\_resolve.cc: reorder the includes ([pr#10505](#), Willem Jan Withagen)
- common: global/signal\_handler: use sig\_str instead of sys\_siglist ([pr#10633](#), John Coyle)
- core,cephfs: Revert "osd/ReplicatedPG: for sync-read it don't calculate OSD op r\_prep..." ([issue#16908](#), [pr#10875](#), Samuel Just)
- core,cephfs: mon/mds: add err info when load\_metadata is abnormal ([pr#10176](#), huanwen ren)
- core,common: osd/OSD.cc: remove unneeded returns ([pr#11043](#), Michal Jarzabek)
- core,pybind: python-rados: extends ReadOp/WriteOp API ([pr#9944](#), Mehdi Abaakouk)
- core,pybind: python-rados: implement new aio\_stat. ([pr#11006](#), Iain Buclaw)
- core,pybind: qa/workunits/rados/test\_python.sh: Allow specifying Python executable ([pr#10782](#), Oleh Prypin)
- core: os/filestore/LFNIndex: remove unused variable 'subdir\_path' ([pr#8959](#), huangjun)
- core: Create ceph-mgr ([pr#10328](#), John Spray, Tim Serong)

- core: FileJournal: Remove obsolete `_check_disk_write_cache` function ([pr#11073](#), Brad Hubbard)
- core: Lua object class support ([pr#7338](#), Noah Watkins)
- core: OSD crash with Hammer to Jewel Upgrade: void `FileStore::init_temp_collections()` ([issue#16672](#), [pr#10565](#), David Zafman)
- core: OSD.cc: remove unneeded return ([pr#9701](#), Michal Jarzabek)
- core: OSD: avoid FileStore finisher deadlock in `osd_lock` when shutdown OSD ([pr#11052](#), Haomai Wang)
- core: ObjectCacher: fix `last_write` check in `bh_write_adjacencies()` ([issue#16610](#), [pr#10304](#), Yan, Zheng)
- core: ReplicatedPG: call `op_applied` for `submit_log_entries` based repops ([pr#9489](#), Samuel Just)
- core: Wip 16998 ([issue#16998](#), [pr#10688](#), Samuel Just)
- core: ceph-create-keys: add missing argument comma ([pr#11123](#), Patrick Donnelly)
- core: ceph-create-keys: fix existing-but-different case ([issue#16255](#), [pr#10415](#), John Spray)
- core: ceph-disk: partprobe should block udev induced BLKRRPART ([issue#15176](#), [pr#9330](#), Marius Vollmer, Loic Dachary)
- core: ceph-disk: timeout ceph-disk to avoid blocking forever ([issue#16580](#), [pr#10262](#), Loic Dachary)
- core: ceph-objectstore-tool: add a way to split filestore directories offline ([issue#17220](#), [pr#10776](#), Josh Durgin)
- core: ceph.in: python 3 compatibility of the ceph CLI ([pr#9702](#), Oleh Prypin)
- core: ceph\_mon: use `readdir()` as `readdir_r()` is deprecated ([pr#11047](#), Kefu Chai)
- core: cephx: Fix multiple segfaults due to attempts to encrypt or decrypt ([issue#16266](#), [pr#9703](#), Brad Hubbard)
- core: <https://github.com/ceph/ceph/pull/11052> ([pr#10371](#), Yan Jun)
- core: include write error codes in the pg log ([issue#14468](#), [pr#10170](#), Josh Durgin)
- core: kv/MemDB: fix assert triggered by `m_total_bytes` underflow ([pr#10471](#), xie xingguo)
- core: kv/RocksDB: add perfcounter for `submit_transaction_sync` operation ([pr#9770](#), Haodong Tang)

- core: logmon: check is\_leader() before doing any work on get\_trim\_to() ([pr#10342](#), song baisen)
- core: memstore: clone zero-fills holes from source range ([pr#11157](#), Casey Bodley)
- core: message: optimization for message priority strategy ([pr#8687](#), yaoning)
- core: messages/MForward: fix encoding features ([issue#17365](#), [pr#11180](#), Sage Weil)
- core: mgr/MgrClient: fix ms\_handle\_reset ([pr#11298](#), Sage Weil)
- core: mgr/MgrMap: initialize all fields ([issue#17492](#), [pr#11308](#), Sage Weil)
- core: mon/ConfigKeyService: pass strings by const ref ([pr#10618](#), Michal Jarzabek)
- core: mon/LogMonitor: move C\_Log struct to cc file ([pr#10721](#), Michal Jarzabek)
- core: mon/MonClient.h: pass strings by const reference ([pr#10605](#), Michal Jarzabek)
- core: mon/MonDBStore: fix assert which never fires ([pr#10706](#), xie xingguo)
- core: mon/MonitorDBStore: do not use snapshot iterator; close on close ([pr#10102](#), Sage Weil)
- core: mon/OSDMonitor.cc: remove use of boost assign ([pr#11060](#), Michal Jarzabek)
- core: mon/PGMonitor: batch filter pg states; add sanity check ([pr#9394](#), xie xingguo)
- core: mon/PGMonitor: calc the %USED of pool using used/(used+avail) ([issue#16933](#), [pr#10584](#), Kefu Chai)
- core: mon/PGMonitor: move C\_Stats struct to cc file ([pr#10719](#), Michal Jarzabek)
- core: mon/PaxosService: make the return value type inconsistent ([pr#10231](#), zhang.zezhu)
- core: mon/osdmonitor: fix incorrect output of "osd df" due to osd out ([issue#16706](#), [pr#10308](#), xie xingguo)
- core: msg/AsyncMessenger: change return type to void ([pr#10230](#), Michal Jarzabek)
- core: msg/Messenger: add const and override to function ([pr#10183](#), Michal Jarzabek)
- core: msg/async/AsyncConnection: replace Mutex with std::mutex for performance ([issue#16714](#), [issue#16715](#), [pr#10340](#), Haomai Wang)
- core: msg/async/Event: ensure not refer to member variable which may destroyed ([issue#16714](#), [pr#10369](#), Haomai Wang)

- core: msg/async/kqueue: avoid remove nonexistent kqueue event ([pr#9869](#), Haomai Wang)
- core: msg/async: Support close idle connection feature ([issue#16366](#), [pr#9783](#), Haomai Wang)
- core: msg/async: allow other async backend implementations ([pr#10264](#), Haomai Wang)
- core: msg/async: avoid set out of range ms\_async\_op\_threads option ([pr#11200](#), Haomai Wang)
- core: msg/async: connect authorizer fix + recv\_buf size ([pr#9784](#), Ilya Dryomov)
- core: msg/async: harden error logic handle ([pr#9781](#), Haomai Wang)
- core: msg/async: remove fd output in log prefix ([pr#11199](#), Haomai Wang)
- core: msg/async: remove file event lock ([issue#16554](#), [issue#16552](#), [pr#10090](#), Haomai Wang)
- core: msg/simple/Pipe: eliminating casts for the comparing of len and recv\_max\_prefetch ([pr#10273](#), zhang.zezhu)
- core: msg/simple: fix wrong condition checking of writing TAG\_CLOSE on closing ([pr#10343](#), xie xingguo)
- core: msg/simple: wait dispatch\_queue until all pipes closed ([issue#16472](#), [pr#9930](#), Haomai Wang)
- core: msg: make async backend default ([pr#10746](#), Haomai Wang)
- core: msg: mark daemons down on RST + ECONNREFUSED ([pr#8558](#), Piotr Dałek)
- core: os/FuseStore: fix several FuseStore issues ([pr#10723](#), Sage Weil)
- core: os/MemStore: move BufferlistObject to .cc file ([pr#10833](#), Michal Jarzabek)
- core: os/ObjectStore: fix return code of collection\_empty() method ([pr#11050](#), xie xingguo)
- core: os/RocksDBStore: use effective Get API instead of iterator api ([pr#9411](#), Jianjian Huo, Haomai Wang, Mark Nelson)
- core: os/filestore/FDCache: fix bug when filestore\_fd\_cache\_shards = 0 ([pr#11048](#), jimifm)
- core: os/filestore/FileJournal: error out if FileJournal is not a file ([issue#17307](#), [pr#11146](#), Kefu Chai)
- core: os/filestore: add sanity checks and cleanups for mount() process ([pr#9734](#), xie xingguo)

- core: os/filestore: disable use of splice by default ([pr#11113](#), Haomai Wang)
- core: osd/OSD.cc: remove repeated searching of map ([pr#10986](#), Michal Jarzabek)
- core: osd/OSD.cc: remove unneeded searching of maps ([pr#11039](#), Michal Jarzabek)
- core: osd/OSD.h: add const to member functions ([pr#11114](#), Michal Jarzabek)
- core: osd/OSD.h: move some members under private ([pr#11121](#), Michal Jarzabek)
- core: osd/OSD.h: remove unneeded line ([pr#8980](#), Michal Jarzabek)
- core: osd/OSDMonitor: misc. cleanups ([pr#10739](#), xie xingguo)
- core: osd/OSDMonitor: misc. fixes ([pr#10491](#), xie xingguo)
- core: osd/ReplicatedBackend: add sanity check during build\_push\_op() ([pr#9491](#), Yan Jun)
- core: osd/ReplicatedPG: for sync-read it don't cacl l\_osd\_op\_r\_prepare\_lat. ([pr#10365](#), Jianpeng Ma)
- core: osd/ReplicatedPG: remove class redeclaration ([pr#11041](#), Michal Jarzabek)
- core: osd/ReplicatedPG: remove unused param "op" from generate\_subop() ([pr#10811](#), jimifm)
- core: osd/Watch: add consts to member functions ([pr#10251](#), Michal Jarzabek)
- core: osd/osd\_type: check if pool is gone during check\_new\_interval() ([pr#10859](#), xie xingguo)
- core: osd/osdmonitor: pool of objects and bytes beyond quota should all be warn ([pr#9085](#), huanwen ren)
- core: osdc/objecter: misc fixes ([pr#10826](#), xie xingguo)
- core: pass string by const ref and add override to virtual function ([pr#9082](#), Michal Jarzabek)
- core: qa/workunits/objectstore/test\_fuse.sh: make test\_fuse.sh work with filestore ([pr#11057](#), Sage Weil)
- core: rados: add option to include clones when doing flush or evict ([pr#9698](#), Mingxin Liu)
- core: subman: use replace instead of format ([issue#16961](#), [pr#10620](#), Loic Dachary)
- core: test/common/Throttle.cc: fix race in shutdown ([pr#10094](#), Samuel Just)
- core: test: add the necessary judgment ([pr#9694](#), huanwen ren)
- core: tox.ini: remove extraneous coverage -omit option ([pr#10943](#), Josh Durgin)

- core: udev: always populate /dev/disk/by-parttypeuuid ([issue#16351](#), [pr#9885](#), Loic Dachary)
- core: os/FuseStore: remove unneeded header file ([pr#10799](#), Michal Jarzabek)
- core: os/MemStore: move OmapIteratorImpl to cc file ([pr#10803](#), Michal Jarzabek)
- core: os/Memstore.h: add override to virtual functions ([pr#10801](#), Michal Jarzabek)
- core: os/Memstore: move PageSetObject class to .cc file ([pr#10817](#), Michal Jarzabek)
- core: os/bluestore: remove unused head file. ([pr#11186](#), Jianpeng Ma)
- core: safe\_io: Improve portability by replacing loff\_t type usage with off\_t. ([pr#9767](#), John Coyle)
- core: src/kv/MemDB.cc: the type of the parameter of push\_back() does not match the ops's value\_type ([pr#10455](#), Willem Jan Withagen)
- core: msg/simple: apply prefetch policy more precisely ([pr#10344](#), xie xingguo)
- core: CompatSet.h: remove unneeded inline ([pr#10071](#), Michal Jarzabek)
- core: Objclass perm feedback ([pr#10313](#), Noah Watkins)
- core: arch/arm.c: remove unnecessary variable read for simplicity ([pr#10821](#), Weibing Zhang)
- crush: don't normalize input of crush\_ln iteratively ([pr#10935](#), Piotr Dałek)
- crush: reset bucket->h.items[i] when removing tree item ([issue#16525](#), [pr#10093](#), Kefu Chai)
- crush: CrushCompiler.cc:884 ([pr#10952](#), xu biao)
- crush: CrushCompiler: error out as long as parse fails ([issue#17306](#), [pr#11144](#), Kefu Chai)
- doc: Add documentation about snapshots ([pr#10436](#), Greg Farnum)
- doc: Add two options to radosgw-admin.rst manpage ([issue#17281](#), [pr#11134](#), Thomas Serlin)
- doc: Changed config parameter "rgw keystone make new tenants" in radosgw multitenancy ([issue#17293](#), [pr#11127](#), SirishaGuduru)
- doc: Modification for "TEST S3 ACCESS" section in "INSTALL CEPH OBJECT GATEWAY" page ([pr#9089](#), la-sguduru)
- doc: Update developer docs for cmake paths ([pr#11163](#), John Spray)

- doc: add “-orphan-stale-secs” to radosgw-admin(8) ([issue#17280](#), [pr#11097](#), Ken Dreyer)
- doc: add \$pid metavar conf doc ([pr#11172](#), Patrick Donnelly)
- doc: add Backporting section to Essentials chapter ([issue#15497](#), [pr#10457](#), Nathan Cutler)
- doc: add Prepare tenant section to Testing in the cloud chapter ([pr#10413](#), Nathan Cutler)
- doc: add Upload logs to archive server section... ([pr#10414](#), Nathan Cutler)
- doc: add client config ref ([issue#16743](#), [pr#10434](#), Patrick Donnelly)
- doc: add graphic for cap bit field ([pr#10897](#), Patrick Donnelly)
- doc: add missing PR to hammer 0.94.8 release notes ([pr#10900](#), Nathan Cutler)
- doc: add openSUSE instructions to quick-start-preflight ([pr#10454](#), Nathan Cutler)
- doc: add rgw\_enable\_usage\_log option in Rados Gateway admin guide ([issue#16604](#), [pr#10159](#), Mike Hackett)
- doc: add troubleshooting steps for ceph-fuse ([pr#10374](#), Ken Dreyer)
- doc: admin/build-doc: bypass sanity check if building doc ([issue#16940](#), [pr#10623](#), Kefu Chai)
- doc: ceph-authtool man page option is -print-key not -print ([pr#9731](#), Brad Hubbard)
- doc: ceph-deploy mon add doesn't take multiple nodes ([pr#10085](#), Chengwei Yang)
- doc: clarify rbd size units ([pr#11303](#), Ilya Dryomov)
- doc: cleanup outdated radosgw description ([pr#11248](#), Jiaying Ren)
- doc: describe libvirt client logging ([pr#10542](#), Ken Dreyer)
- doc: do not list all major versions in get-packages.rst ([pr#10899](#), Nathan Cutler)
- doc: doc/cephfs: explain the various health messages ([pr#10244](#), John Spray)
- doc: doc/dev: Fix missing code section due to no lexer for “none” ([pr#9083](#), Brad Hubbard)
- doc: doc/radosgw: fix description of response elements ‘Part’ ([pr#10641](#), weiqiaomiao)
- doc: doc/radosgw: rename config.rst to config-fcgi.rst ([pr#10381](#), Nathan Cutler)
- doc: extend the CephFS troubleshooting guide ([pr#10458](#), Greg Farnum)

- doc: fix broken link in SHEC erasure code plugin ([issue#16996](#), [pr#10675](#), Albert Tu)
- doc: fix description for rsize and rasize ([pr#11101](#), Andreas Gerstmayr)
- doc: fix rados/configuration/osd-config-ref.rst ([pr#10619](#), Chengwei Yang)
- doc: fix singleton example in Developer Guide ([pr#10830](#), Nathan Cutler)
- doc: fix some nits in release notes and releases table ([pr#10903](#), Nathan Cutler)
- doc: fix standby replay config ([issue#16664](#), [pr#10268](#), Patrick Donnelly)
- doc: fix wrong osdkeepalive name in mount.ceph manpage ([pr#10840](#), Zhi Zhang)
- doc: fix/add changelog for 10.2.2, 0.94.7, 0.94.8 ([pr#10895](#), Sage Weil)
- doc: format 2 now is the default image format ([pr#10705](#), Chengwei Yang)
- doc: lgtm (build verified f23) ([pr#9745](#), weiqiaomiao)
- doc: mailmap updates for upcoming 11.0.0 ([pr#9301](#), Yann Dupont)
- doc: manual instructions to set up mds daemon ([pr#11115](#), Peter Maloney)
- doc: missing “make vstart” in quick\_guide.rst ([pr#11226](#), Leo Zhang)
- doc: more details for pool deletion ([pr#10190](#), Ken Dreyer)
- doc: peering.rst, fix typo ([pr#10131](#), Brad Hubbard)
- doc: perf\_counters.rst fix trivial typo ([pr#10292](#), Brad Hubbard)
- doc: rbdmap: specify bash shell interpreter ([issue#16608](#), [pr#10733](#), Jason Dillaman)
- doc: release-notes.rst: draft 0.94.8 release notes ([pr#10730](#), Nathan Cutler)
- doc: remove btrfs contradiction ([pr#9758](#), Nathan Cutler)
- doc: remove i386 from minimal hardware recommendations ([pr#10276](#), Kefu Chai)
- doc: remove old references to inktank premium support ([pr#11182](#), Alfredo Deza)
- doc: remove the description of deleted options ([issue#17041](#), [pr#10741](#), MinSheng Lin)
- doc: rgw, doc: fix formatting around Keystone-related options. ([pr#10331](#), Radoslaw Zarzynski)
- doc: rgw/doc: fix indent ([pr#10676](#), Yan Jun)
- doc: rm SysV instructions, add systemd ([pr#10184](#), Ken Dreyer)

- doc: silence sphinx warnings ([pr#10621](#), Kefu Chai)
- doc: small standby doc edits ([pr#10479](#), Patrick Donnelly)
- doc: update CephFS “early adopters” info ([pr#10068](#), John Spray)
- doc: update canonical tarballs URL ([pr#9695](#), Ken Dreyer)
- doc: update rbd glance configuration notes ([pr#10629](#), Jason Dillaman)
- doc: update s3 static webiste feature support status ([pr#10223](#), Jiaying Ren)
- doc: changelog: add v10.2.3 ([pr#11238](#), Abhishek Lekshmanan)
- doc: install: Use <https://> for download.ceph.com ([pr#10709](#), Colin Walters)
- doc: release-notes: v0.94.9 ([pr#10927](#), Sage Weil)
- doc: release-notes: v10.2.3 jewel ([pr#11234](#), Abhishek Lekshmanan)
- doc: Add UK mirror and update copyright ([pr#10531](#), Patrick McGarry)
- doc: README.md: replace package build instructions with tarball instructions ([pr#10829](#), Sage Weil)
- doc: Removed reference about pool ownership based on BZ#1368528 ([pr#11063](#), Bara Ancincova)
- librados: use bufferlist instead of buffer::list in public header ([pr#10632](#), Ryne Li)
- librados: Rados-stripper: Flexible string matching for not found attributes ([pr#10577](#), Willem Jan Withagen)
- librados: librados examples: link and include from current source tree by default. ([issue#15100](#), [pr#8189](#), Jesse Williamson)
- librbd: API methods to directly acquire and release the exclusive lock ([issue#15632](#), [pr#9592](#), Mykola Golub)
- librbd: add consistency groups operations with images ([pr#10034](#), Victor Denisov)
- librbd: add explicit shrink check while resizing images ([pr#9878](#), Vaibhav Bhembre)
- librbd: asynchronous v2 image creation ([issue#15321](#), [pr#9585](#), Venky Shankar)
- librbd: backward/forward compatibility for update\_features ([issue#17330](#), [pr#11155](#), Jason Dillaman)
- librbd: block name prefix might overflow fixed size C-string ([issue#17310](#), [pr#11148](#), Jason Dillaman)

- librbd: cache was not switching to writeback after first flush ([issue#16654](#), [pr#10762](#), Jason Dillaman)
- librbd: corrected use-after-free in Imagewatcher ([issue#17289](#), [pr#11112](#), Jason Dillaman)
- librbd: deadlock when replaying journal during image open ([issue#17188](#), [pr#10945](#), Jason Dillaman)
- librbd: delay acquiring lock if image watch has failed ([issue#16923](#), [pr#10574](#), Jason Dillaman)
- librbd: discard hangs when 'rbd\_skip\_partial\_discard' is enabled ([issue#16386](#), [pr#10060](#), Mykola Golub)
- librbd: extract group module from librbd/internal ([pr#11070](#), Victor Denisov)
- librbd: failed assertion after shrinking a clone image twice ([issue#16561](#), [pr#10072](#), Jason Dillaman)
- librbd: fix missing return statement if failed to get mirror image state ([pr#10136](#), runsisi)
- librbd: fix possible inconsistent state when disabling mirroring fails ([issue#16984](#), [pr#10711](#), Jason Dillaman)
- librbd: ignore partial refresh error when acquiring exclusive lock ([issue#17227](#), [pr#11044](#), Jason Dillaman)
- librbd: initial hooks for client-side, image-extent cache in IO path ([pr#9121](#), Jason Dillaman)
- librbd: interlock image refresh and exclusive lock operations ([issue#16773](#), [issue#17015](#), [pr#10770](#), Jason Dillaman)
- librbd: memory leak in MirroringWatcher::notify\_image\_updated ([pr#11306](#), Mykola Golub)
- librbd: optimize away unnecessary object map updates ([issue#16707](#), [issue#16689](#), [pr#10332](#), Jason Dillaman)
- librbd: optionally unregister "laggy" journal clients ([issue#14738](#), [pr#10378](#), Mykola Golub)
- librbd: permit disabling journaling if in corrupt state ([issue#16740](#), [pr#10712](#), Jason Dillaman)
- librbd: possible deadlock if cluster connection closed after image ([issue#17254](#), [pr#11037](#), Jason Dillaman)
- librbd: potential deadlock closing image with in-flight readahead ([issue#17198](#),

- pr#11152, Jason Dillaman)
- librbd: potential double-unwatch of watch handle upon error ([issue#17210](#), pr#10974, Jason Dillaman)
  - librbd: potential seg fault when blacklisting an image client ([issue#17251](#), pr#11034, Jason Dillaman)
  - librbd: prevent creation of clone from non-primary mirrored image ([issue#16449](#), pr#10123, Mykola Golub)
  - librbd: prevent creation of v2 image ids that are too large ([issue#16887](#), pr#10581, Jason Dillaman)
  - mds: Add path filtering for dump cache ([issue#11171](#), pr#9925, Douglas Fuller)
  - mds: Kill C\_SaferCond in evict\_sessions() ([issue#16288](#), pr#9971, Douglas Fuller)
  - mds: Return “committing” rather than “committed” member in get\_committing ([pr#10250](#), Greg Farnum)
  - mds: Set mds\_snap\_max\_uid to 4294967294 ([pr#11016](#), Wido den Hollander)
  - mds: add assertion in handle\_slave\_rename\_prep ([issue#16807](#), pr#10429, John Spray)
  - mds: add assertions for standby\_daemons invariant ([issue#16592](#), pr#10316, Patrick Donnelly)
  - mds: add health warning for oversized cache ([issue#16570](#), pr#10245, John Spray)
  - mds: add maximum fragment size constraint ([issue#16164](#), pr#9789, Patrick Donnelly)
  - mds: add perf counters for MDLog replay and SessionMap ([pr#10539](#), John Spray)
  - mds: catch duplicates in DamageTable ([issue#17173](#), pr#11137, John Spray)
  - mds: fix Session::check\_access() ([issue#16358](#), pr#9769, Yan, Zheng)
  - mds: fix daemon selection when starting ranks ([pr#10540](#), John Spray)
  - mds: fix shutting down mds timed-out due to deadlock ([issue#16396](#), pr#9884, Zhi Zhang)
  - mds: fix up \_dispatch ref-counting semantics ([pr#10533](#), Greg Farnum)
  - mds: fixup dump Formatter' type error; add path\_ino and is\_primary in the CDentry::dump() ([pr#10119](#), huanwen ren)
  - mds: handle blacklisting during journal recovery ([issue#17236](#), pr#11138, John Spray)

- mds: log path with CDir damage messages ([issue#16973](#), [pr#10996](#), John Spray)
- mds: move Finisher to unlocked shutdown ([issue#16042](#), [pr#10142](#), Patrick Donnelly)
- mds: populate DamageTable from scrub and log more quietly ([issue#16016](#), [pr#11136](#), John Spray)
- mds: remove fail-safe queueing replay request ([issue#17271](#), [pr#11078](#), Yan, Zheng)
- mds: remove max\_mds config option ([issue#17105](#), [pr#10914](#), Patrick Donnelly)
- mds: remove unused MDSDaemon::objecter ([pr#10566](#), Patrick Donnelly)
- mds: snap failover fixes ([pr#9955](#), Yan, Zheng)
- mds: trim null dentries proactively ([issue#16919](#), [pr#10606](#), John Spray)
- mds: unuse Class and cleanup ([pr#10399](#), huanwen ren)
- mds: use reference to avoid copy ([pr#10191](#), Patrick Donnelly)
- mds: MDCache.h: remove unneeded access specifier ([pr#10901](#), Michal Jarzabek)
- mds: MDSDaemon: move C\_MDS\_Tick class to .cc file ([pr#11220](#), Michal Jarzabek)
- mgr: implement con reset handling ([pr#11299](#), Sage Weil)
- mgr: squash compiler warnings ([pr#11307](#), John Spray)
- mon: MonClient may hang on pinging an unresponsive monitor ([pr#9259](#), xie xingguo)
- mon: Monitor: validate prefix on handle\_command() ([issue#16297](#), [pr#9700](#), You Ji)
- mon: OSDMonitor: Missing nearfull flag set ([pr#11082](#), Igor Podoski)
- mon: change osdmap flags set and unset messages ([issue#15983](#), [pr#9252](#), Vikhyat Umrao)
- mon: clear list in better way ([pr#9718](#), song baisen)
- mon: do not recalculate 'to\_remove' when it's known ([pr#9717](#), song baisen)
- mon: misc cleanups ([pr#10591](#), xie xingguo)
- mon: remove the redundant cancel\_probe\_timeout function ([pr#10261](#), song baisen)
- mon: remove the redundant is\_active judge in PaxosService ([pr#9749](#), song baisen)
- mon: tear down standby replays on MDS rank stop ([issue#16909](#), [pr#10628](#), John Spray)
- mon: use clearer code structure ([pr#10192](#), Patrick Donnelly)
- mon: validate states transmitted in beacons ([issue#16592](#), [pr#10428](#), John Spray)

- mon: wait 10m (not 5m) before marking down OSDs out ([pr#11184](#), Sage Weil)
- mon: write fsid use the right return value ([pr#10197](#), song baisen)
- mon: Elector:move C\_ElectionExpire class to cc file ([pr#10416](#), Michal Jarzabek)
- mon: HealthMonitor: add override to virtual functs ([pr#10549](#), Michal Jarzabek)
- mon: HealthMonitor: remove unneeded include ([pr#10563](#), Michal Jarzabek)
- mon: MonClient.h: delete copy constr and assing op ([pr#10599](#), Michal Jarzabek)
- mon: MonClient: move C\_CancelMonCommand to cc file ([pr#10392](#), Michal Jarzabek)
- mon: MonClient: move C\_Tick struct to cc file ([pr#10383](#), Michal Jarzabek)
- mon: Monitor.h: add override to virtual functions ([pr#10515](#), Michal Jarzabek)
- mon: Monitor: move C\_Scrub, C\_ScrubTimeout to .cc ([pr#10513](#), Michal Jarzabek)
- mon: OSDMonitor.cc: remove unneeded casts ([pr#10575](#), Michal Jarzabek)
- mon: Paxos: move classes to .cc file ([pr#11215](#), Michal Jarzabek)
- mon: PaxosService: move classes to cc file ([pr#10529](#), Michal Jarzabek)
- mon: remove the redundant list swap in paxos commit\_proposal ([pr#10011](#), song baisen)
- msgr: set close on exec flag ([issue#16390](#), [pr#9772](#), Kefu Chai)
- msgr: Acceptor.h: add override to virtual function ([pr#10422](#), Michal Jarzabek)
- msgr: Acceptor: move include to cc file ([pr#10441](#), Michal Jarzabek)
- msgr: AsyncConnection: add const to mem functions ([pr#10302](#), Michal Jarzabek)
- msgr: AsyncMessenger.cc: remove unneeded cast ([pr#10141](#), Michal Jarzabek)
- msgr: AsyncMessenger: add const to function ([pr#10114](#), Michal Jarzabek)
- msgr: AsyncMessenger: move C\_handle\_reap class to cc ([pr#10113](#), Michal Jarzabek)
- msgr: AsyncMessenger: move C\_processor\_accept class ([pr#9991](#), Michal Jarzabek)
- msgr: AsyncMessenger: remove unneeded include file ([pr#10195](#), Michal Jarzabek)
- msgr: AsyncMessenger: remove unused function ([pr#10163](#), Michal Jarzabek)
- msgr: EventKqueue.h: add override to virtual func ([pr#10318](#), Michal Jarzabek)
- msgr: EventPoll.h: add override to virtual functions ([pr#10314](#), Michal Jarzabek)
- msgr: EventSelect.h: add override to virtual funct ([pr#10321](#), Michal Jarzabek)

- msgr: EventSelect: move includes to cc file ([pr#10333](#), Michal Jarzabek)
- msgr: FastStrategy.h: add override to virtual funct ([pr#10482](#), Michal Jarzabek)
- msgr: Message.h: add const to member function ([pr#10354](#), Michal Jarzabek)
- msgr: Message.h: remove code duplication ([pr#10356](#), Michal Jarzabek)
- msgr: QueueStrategy: add override to virtual functs ([pr#10503](#), Michal Jarzabek)
- msgr: Stack.h: delete copy constr and assign op ([pr#11107](#), Michal Jarzabek)
- msgr: async/Event.h: add const to member function ([pr#10224](#), Michal Jarzabek)
- msgr: async: remove unused code. ([pr#11247](#), Jianpeng Ma)
- osd: bail out if transaction size overflows ([issue#16982](#), [pr#10753](#), Kefu Chai)
- osd: cleanup options and other redundancies ([pr#10450](#), xie xingguo)
- osd: drop unused variables/methods ([pr#10559](#), xie xingguo)
- osd: fix the mem leak of RepGather ([issue#16801](#), [pr#10423](#), Kefu Chai)
- osd: fixups to explicitly persistently missing sets ([pr#10405](#), Samuel Just)
- osd: increment stats on recovery pull also ([issue#16277](#), [pr#10152](#), Kefu Chai)
- osd: limit omap data in push op ([issue#16128](#), [pr#9894](#), Wanlong Gao)
- osd: minor performance improvements ([pr#10470](#), xie xingguo)
- osd: minor performance improvements and fixes ([pr#10526](#), xie xingguo)
- osd: misc fixes and cleanups ([pr#10610](#), xie xingguo)
- osd: miscellaneous fixes ([pr#10572](#), xie xingguo)
- osd: more cleanups ([pr#10548](#), xie xingguo)
- osd: object class loading and execution permissions ([pr#9972](#), Noah Watkins)
- osd: pass shared\_ptr by const reference ([pr#11266](#), Michal Jarzabek)
- osd: persist the missing set explicitly ([pr#10334](#), Samuel Just)
- osd: remove dispatch queue check since we don't queue hb message to this ([pr#9947](#), Haomai Wang)
- osd: remove duplicated function ([pr#9117](#), Wei Jin)
- osd: replace ceph::atomic\_t with std::atomic in osd module. ([pr#9138](#), Xiaowei Chen)

- osd: should not look up an empty pg ([issue#17380](#), [pr#11208](#), Kefu Chai, Loic Dachary)
- osd: small cleanups ([pr#9980](#), Wanlong Gao)
- osd: subscribe for old osdmmaps when pause flag is set ([issue#17023](#), [pr#10725](#), Kefu Chai)
- osd:preserve allocation hint attribute during recovery ([pr#9452](#), yaoning)
- osd: osd-fast-mark-down.sh: fix typo in variable assignments ([pr#11224](#), Willem Jan Withagen)
- osd: OSD.cc: initialise variable at definition ([pr#11099](#), Michal Jarzabek)
- osd: OSD.cc: remove unneeded searching of map ([pr#11000](#), Michal Jarzabek)
- osd: OSD.h: make some members private ([pr#11085](#), Michal Jarzabek)
- osd: PG.cc: remove unneeded use of count ([pr#11228](#), Michal Jarzabek)
- osd: PGBackend.h: move structs to .cc file ([pr#10975](#), Michal Jarzabek)
- osd: ReplicatedBackend: move classes to cc file ([pr#10967](#), Michal Jarzabek)
- osd: ReplicatedPG.h: add override to virtual funct ([pr#11271](#), Michal Jarzabek)
- osd: ReplicatedPG: move classes to .cc file ([pr#10971](#), Michal Jarzabek)
- osd: ReplicatedPG:move C OSD OnApplied class to cc ([pr#11288](#), Michal Jarzabek)
- osd: Watch.h: remove unneeded forward declaration ([pr#10269](#), Michal Jarzabek)
- osd: osdc/ObjectCacher.h: add const to member functions ([pr#9569](#), Michal Jarzabek)
- osd: osdc/ObjectCacher.h: add const to member functions ([pr#9652](#), Michal Jarzabek)
- osd: osdc/Objecter: move RequestStateHook class to .cc ([pr#10734](#), Michal Jarzabek)
- pybind: Port Python-based tests and remaining Python bindings to Python 3 ([pr#10177](#), Oleh Prypin)
- pybind: Rework cephfs/setup.py for PyPI ([pr#10315](#), Anirudha Bose)
- pybind: Rework rbd/setup.py for PyPI ([issue#16940](#), [pr#10376](#), Anirudha Bose)
- pybind: global/signal\_handler: dump cmdline instead of arg[0] ([pr#10345](#), Kefu Chai)
- pybind: make rados ready for PyPI ([pr#9833](#), Anirudha Bose)

- pybind: pybind/ceph\_argparse: handle non ascii unicode args ([issue#12287](#), [pr#8943](#), Kefu Chai)
- pybind: Python 3 compatibility for workunits ([pr#10815](#), Anirudha Bose)
- rbd: Allow user to remove snapshot with -force to auto flatten children ([pr#10087](#), Dongsheng Yang)
- rbd: Reviewed-off-by: Ilya Dryomov <[idryomov@gmail.com](mailto:idryomov@gmail.com)> ([issue#16171](#), [pr#10481](#), Jason Dillaman)
- rbd: Reviewed-off-by: Ilya Dryomov <[idryomov@gmail.com](mailto:idryomov@gmail.com)> ([issue#17030](#), [pr#10735](#), Jason Dillaman)
- rbd: bench io-size should not be larger than image size ([issue#16967](#), [pr#10708](#), Jason Dillaman)
- rbd: cleanup - Proxied operations shouldn't result in error messages if replayed ([issue#16130](#), [pr#9724](#), Vikhyat Umrao)
- rbd: cls\_rbd: clean up status from rbd-mirror if image removed ([pr#11142](#), Huan Zhang)
- rbd: cls\_rbd: set omap values in batch during image creation ([pr#9981](#), Dongsheng Yang)
- rbd: inherit the parent image features when cloning an image ([issue#15388](#), [pr#9334](#), Dongsheng Yang)
- rbd: journal: ensure in-flight ops are complete destroying journaler ([issue#17446](#), [pr#11257](#), Mykola Golub, Jason Dillaman)
- rbd: journal: increase concurrency/parallelism of journal recorder ([issue#15259](#), [pr#10445](#), Ricardo Dias)
- rbd: journal: move JournalTrimmer::C\_RemoveSet struct ([pr#10912](#), Michal Jarzabek)
- rbd: qa/workunits/rbd: before removing image make sure it is not bootstrapped ([issue#16555](#), [pr#10155](#), Mykola Golub)
- rbd: qa/workunits/rbd: check status also in pool dir after asok commands ([pr#11291](#), Mykola Golub)
- rbd: qa/workunits/rbd: set image-meta on primary image and wait it is replicated ([pr#11294](#), Mykola Golub)
- rbd: qa/workunits/rbd: small fixup and improvements for rbd-mirror tests ([pr#10483](#), Mykola Golub)
- rbd: qa/workunits/rbd: wait for image deleted before checking health ([pr#10545](#), Mykola Golub)

- rbd: qa/workunits: support filtering cls\_rbd unit test cases ([issue#16529](#), [pr#10714](#), Jason Dillaman)
- rbd: rbd-mirror: 'wait\_for\_scheduled\_deletion' callback might deadlock ([issue#16491](#), [pr#9964](#), Jason Dillaman)
- rbd: rbd-mirror: Add sparse read for sync image ([issue#16780](#), [pr#11005](#), tianqing)
- rbd: rbd-mirror: add additional test scenarios ([pr#10488](#), lande1234)
- rbd: rbd-mirror: concurrent access of event might result in heap corruption ([issue#17283](#), [pr#11104](#), Jason Dillaman)
- rbd: rbd-mirror: force-promoted image will remain R/O until rbd-mirror daemon restarted ([issue#16974](#), [pr#11090](#), Jason Dillaman)
- rbd: rbd-mirror: gracefully fail if object map is unavailable ([issue#16558](#), [pr#10065](#), Jason Dillaman)
- rbd: rbd-mirror: gracefully handle being blacklisted ([issue#16349](#), [pr#9970](#), Jason Dillaman)
- rbd: rbd-mirror: image deleter should use pool id + global image uuid for key ([issue#16538](#), [issue#16227](#), [pr#10484](#), Jason Dillaman)
- rbd: rbd-mirror: improve split-brain detection logic ([issue#16855](#), [pr#10703](#), Jason Dillaman)
- rbd: rbd-mirror: include local pool id in resync throttle unique key ([issue#16536](#), [pr#10254](#), Jason Dillaman)
- rbd: rbd-mirror: non-primary image is recording journal events during image sync ([pr#10462](#), Jason Dillaman)
- rbd: rbd-mirror: potential IO stall when using asok flush request ([issue#16708](#), [pr#10432](#), Jason Dillaman)
- rbd: rbd-mirror: potential assertion failure during error-induced shutdown ([issue#16956](#), [pr#10613](#), Jason Dillaman)
- rbd: rbd-mirror: potential race condition during failure shutdown ([issue#16980](#), [pr#10667](#), Jason Dillaman)
- rbd: rbd-mirror: quiesce in-flight event commits before shut down ([issue#17355](#), [pr#11185](#), Jason Dillaman)
- rbd: rbd-mirror: reduce memory footprint during journal replay ([issue#16223](#), [pr#10341](#), Jason Dillaman)
- rbd: rbd-mirror: remove ceph\_test\_rbd\_mirror\_image\_replay test case ([issue#16539](#), [pr#10083](#), Mykola Golub)

- rbd: rbd-mirror: replaying state should include flush action ([issue#16970](#), [pr#10627](#), Jason Dillaman)
- rbd: rbd-mirror: replicate dynamic feature updates ([issue#16213](#), [pr#10980](#), Mykola Golub)
- rbd: rbd-mirror: replicate image metadata settings ([issue#16212](#), [pr#11168](#), Mykola Golub)
- rbd: rbd-mirror: snap rename does not properly replicate to peers ([issue#16622](#), [pr#10249](#), Jason Dillaman)
- rbd: rbd-nbd does not properly handle resize notifications ([issue#15715](#), [pr#9291](#), Mykola Golub)
- rbd: rbd-nbd: fix kernel deadlock during teuthology testing ([issue#16921](#), [pr#10985](#), Jason Dillaman)
- rbd: recognize lock\_on\_read option ([pr#11313](#), Ilya Dryomov)
- rbd: return error if we specified a wrong image name for rbd du ([issue#16987](#), [pr#11031](#), Dongsheng Yang)
- rbd: test/librbd/fsx: enable exclusive-lock feature in krbd mode ([pr#10984](#), Ilya Dryomov)
- rbd: test/rbd: fix possible mock journal race conditions ([issue#17317](#), [pr#11153](#), Jason Dillaman)
- rbd: test: cmake related fixups for rbd tests ([pr#10124](#), Mykola Golub)
- rbd: test: run-rbd-tests test cmake fixup ([pr#10134](#), Mykola Golub)
- rbd: test: use wrapper that respects RBD\_FEATURES when creating rbd image ([issue#16717](#), [pr#10348](#), Mykola Golub)
- rbd: udev: add krbd readahead placeholder ([pr#10841](#), Nick Fisk)
- rbd: rbd\_mirror/ImageSyncThrottler: move struct to .cc ([pr#10928](#), Michal Jarzabek)
- rgw: (build verified, provably unused/not aliased) ([pr#9993](#), weiqiaomiao)
- rgw: Add documentation for the Multi-tenancy feature ([pr#9570](#), Pete Zaitcev)
- rgw: Clean up lifecycle thread ([pr#10480](#), Daniel Gryniewicz)
- rgw: Do not archive metadata by default ([issue#17256](#), [pr#11051](#), Pavan Rallabhandi)
- rgw: Fix Host->bucket fallback logic inversion ([issue#15975](#), [issue#17136](#), [pr#10873](#), Robin H. Johnson)

- rgw: Fix for using port 443 with pre-signed urls. ([issue#16548](#), [pr#10088](#), Pritha Srivastava)
- rgw: Fix incorrect content length and range for zero sized objects during range requests ([issue#16388](#), [pr#10207](#), Pavan Rallabhandi)
- rgw: Got rid of recursive mutex. ([pr#10562](#), Adam Kupczyk)
- rgw: RGW : setting socket backlog for via ceph.conf ([issue#16406](#), [pr#9891](#), Feng Guo)
- rgw: RGWMetaSyncCR holds refs to stacks instead of crs ([issue#16666](#), [pr#10301](#), Casey Bodley)
- rgw: Reviewed by: Pritha Srivastava <[prsriwas@redhat.com](mailto:prsriwas@redhat.com)> ([issue#16188](#), [pr#9584](#), Albert Tu)
- rgw: Rgw lifecycle testing ([pr#11131](#), Daniel Gryniewicz)
- rgw: Rgw nfs 28 ([pr#10611](#), Matt Benjamin)
- rgw: add configurables for {data,meta} sync error injection ([pr#10388](#), Yehuda Sadeh)
- rgw: add deadlock detection to RGWCoroutinesManager::run() ([pr#10032](#), Casey Bodley)
- rgw: add lc\_pool when decode or encode struct RGWZoneParams ([pr#10439](#), weiqiaomiao)
- rgw: add missing master\_zone when running with old default region config ([issue#16627](#), [pr#10205](#), Orit Wasserman)
- rgw: add pg\_ver to tombstone\_cache ([pr#9851](#), Casey Bodley)
- rgw: add reinit/rebind logic (ldap) ([pr#10532](#), Matt Benjamin)
- rgw: add return value checking to avoid possible subsequent parse exception ([pr#10241](#), Yan Jun)
- rgw: add suport for Swift-at-root dependent features of Swift API ([issue#16673](#), [pr#10280](#), Pritha Srivastava, Radoslaw Zarzynski)
- rgw: add support for Static Website of Swift API ([pr#9844](#), Radoslaw Zarzynski)
- rgw: add tenant support to multisite sync ([issue#16469](#), [pr#10075](#), Casey Bodley)
- rgw: back off bucket sync on failures, don't store marker ([issue#16742](#), [pr#10355](#), Yehuda Sadeh)
- rgw: better error message when user has no bucket created yet ([issue#16444](#), [pr#10162](#), Gaurav Kumar Garg)

- rgw: clean-up in the authentication infrastructure ([pr#10212](#), Radoslaw Zarzynski)
- rgw: clear realm watch on failed watch\_restart ([issue#16817](#), [pr#10446](#), Casey Bodley)
- rgw: collect skips a specific coroutine stack ([issue#16665](#), [pr#10274](#), Yehuda Sadeh)
- rgw: cosmetic changes only-build verified, f23 ([pr#9931](#), Yan Jun)
- rgw: delete region map after upgrade to zonegroup map ([issue#17051](#), [pr#10831](#), Casey Bodley)
- rgw: do not try to encode or decode time\_t and fix compiling warnings ([pr#10751](#), Kefu Chai)
- rgw: don't fail if lost race when setting acls ([issue#16930](#), [pr#11286](#), Yehuda Sadeh)
- rgw: drop create\_bucket in fwd\_request log message ([pr#10214](#), Abhishek Lekshmanan)
- rgw: eradicate dynamic memory allocation in RGWPostObj. ([pr#11054](#), Radoslaw Zarzynski)
- rgw: file setattr ([pr#8618](#), Matt Benjamin)
- rgw: finish error\_repo cr in stop\_spawned\_services() ([issue#16530](#), [pr#10031](#), Yehuda Sadeh)
- rgw: fix RGWAccessControlPolicy\_SWIFT::create return value check error ([issue#17090](#), [pr#10727](#), weiqiaomiao)
- rgw: fix compilation ([pr#10252](#), Josh Durgin)
- rgw: fix decoding of creation\_time and last\_update. ([issue#17167](#), [pr#11132](#), Orit Wasserman)
- rgw: fix error\_repo segfault in data sync ([issue#16603](#), [pr#10157](#), Casey Bodley)
- rgw: fix failed to create bucket if a non-master zonegroup has a single zone ([pr#10991](#), weiqiaomiao)
- rgw: fix flush\_read\_list() error msg ([pr#10749](#), Jiaying Ren)
- rgw: fix for issue 16494 ([issue#16494](#), [pr#10077](#), Yehuda Sadeh)
- rgw: fix for s3tests failure when ldap auth is not applied ([pr#10669](#), Casey Bodley)
- rgw: fix get object instance returned NoSuchKey error ([issue#17111](#), [pr#10820](#), Yang Honggang)

- rgw: fix is\_admin handling in RGWLDAEngine and introduce acct\_privilege\_t ([pr#10687](#), Radoslaw Zarzynski)
- rgw: fix issue 16435 ([issue#16435](#), [pr#10193](#), Yehuda Sadeh)
- rgw: fix multi-delete query param parsing. ([issue#16618](#), [pr#10187](#), Robin H. Johnson)
- rgw: fix period update -commit return error ([issue#17110](#), [pr#10836](#), weiqiaomiao)
- rgw: fix radosgw daemon core when reopen logs ([issue#17036](#), [pr#10737](#), weiqiaomiao)
- rgw: fix regression with handling double underscore ([issue#16856](#), [pr#10939](#), Orit Wasserman)
- rgw: fix rgw\_bucket\_dir\_entry decode v ([pr#10918](#), Tianshan Qu)
- rgw: fix the error return variable in log message and cleanups ([pr#10138](#), Yan Jun)
- rgw: fix the missing return value ([pr#10122](#), Yan Jun)
- rgw: fix upgrade from old multisite to new multisite configuration ([issue#16751](#), [pr#10368](#), Orit Wasserman)
- rgw: fix wrong variable definition in cls\_version\_check func ([pr#10233](#), weiqiaomiao)
- rgw: fix wrong variable definition in rgw\_cls\_lc\_set\_entry function ([pr#10408](#), weiqiaomiao)
- rgw: for the create\_bucket api, if the input creation\_time is zero, we should set it to 'now' ([issue#16597](#), [pr#10118](#), weiqiaomiao)
- rgw: kill a compile warning for rgw\_sync ([pr#10425](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: lgtm ([pr#9941](#), weiqiaomiao)
- rgw: lgtm (build verified, f23) ([pr#9754](#), John Coyle)
- rgw: lgtm, build verified f23 ([pr#10035](#), Yan Jun)
- rgw: lgtm-build verified, f23 ([pr#10002](#), Yan Jun)
- rgw: lgtm-build verified, f23 ([pr#9985](#), Yan Jun)
- rgw: lgtm-should backport ([pr#9979](#), Yan Jun)
- rgw: log mp upload failures due to parts mismatch ([pr#10424](#), Abhishek Lekshmanan)

- rgw: merge setting flags operation together and cleanups ([pr#10203](#), Yan Jun)
- rgw: miscellaneous cleanups ([pr#10299](#), Yan Jun)
- rgw: multiple fixes for Swift's object expiration ([issue#16705](#), [issue#16684](#), [pr#10330](#), Radoslaw Zarzynski)
- rgw: need to 'open\_object\_section' before dump stats in 'RGWGetUsage...' ([issue#17499](#), [pr#11325](#), weiqiaomiao)
- rgw: obsolete 'radosgw-admin period prepare' command ([issue#17387](#), [pr#11278](#), Gaurav Kumar Garg)
- rgw: radosgw-admin: add “-orphan-stale-secs” to -help ([issue#17280](#), [pr#11098](#), Ken Dreyer)
- rgw: radosgw-admin: zone[group] modify can change realm id ([issue#16839](#), [pr#10477](#), Casey Bodley)
- rgw: raise log levels for common radosgw-admin errors ([issue#16935](#), [pr#10602](#), Shilpa Jagannath)
- rgw: register the correct handler for cls\_user\_complete\_stats ([issue#16624](#), [pr#10151](#), Orit Wasserman)
- rgw: remove bucket index objects when deleting the bucket ([issue#16412](#), [pr#10120](#), Orit Wasserman)
- rgw: remove possible duplicate setting ([pr#10110](#), Yan Jun)
- rgw: remove the field ret from class RGWPutLC ([pr#10726](#), weiqiaomiao)
- rgw: remove unused bufferlist variable ([pr#10194](#), Yan Jun)
- rgw: remove unused realm from radosgw-admin zone modify ([issue#16632](#), [pr#10211](#), Orit Wasserman)
- rgw: remove unused variables ([pr#10589](#), Yan Jun)
- rgw: return “NoSuchLifecycleConfiguration” if lifecycle config does not exist ([pr#10442](#), weiqiaomiao)
- rgw: revert a commit that broke s3 signature validation ([issue#17279](#), [pr#11102](#), Casey Bodley)
- rgw: rgw file: remove busy-wait in RGWLibFS::gc() ([pr#10638](#), Matt Benjamin)
- rgw: rgw ldap: protect rgw::from\_base64 from non-base64 input ([pr#10777](#), Matt Benjamin)
- rgw: rgw ldap: enforce simple\_bind w/LDAPv3 ([pr#10593](#), Matt Benjamin)

- rgw: rgw multisite: RGWCoroutinesManager::run returns status of last cr ([issue#17047](#), [pr#10778](#), Casey Bodley)
- rgw: rgw multisite: RGWDataSyncCR fails on errors from RGWListBucketIndexesCR ([issue#17073](#), [pr#10779](#), Casey Bodley)
- rgw: rgw multisite: fix for assertion in RGWMetaSyncCR ([issue#17044](#), [pr#10743](#), Casey Bodley)
- rgw: rgw multisite: fixes for period puller ([issue#16939](#), [pr#10596](#), Casey Bodley)
- rgw: rgw multisite: trim data logs as peer zones catch up ([pr#10372](#), Casey Bodley)
- rgw: rgw nfs v3 completions ([pr#10745](#), Matt Benjamin)
- rgw: rgw-admin: allow unsetting user's email ([issue#13286](#), [pr#11340](#), Yehuda Sadeh, Weijun Duan)
- rgw: rgw/admin: fix some return values and indents ([pr#9170](#), Yan Jun)
- rgw: rgw/rados: remove confused error printout ([pr#9351](#), Yan Jun)
- rgw: rgw/rgw\_common.cc: modify the end check in RGWHTTPArgs::sys\_get ([pr#9136](#), zhao kun)
- rgw: rgw/rgw\_lc.cc: fix sleep time according to the error message ([pr#10930](#), Weibing Zhang)
- rgw: rgw/rgw\_main: fix unnecessary variables defined ([pr#10475](#), zhang.zezhu)
- rgw: rgw/swift: remove redundant assignment operation ([pr#11292](#), Yan Jun)
- rgw: rgw\_file: pre-assign times ([issue#17367](#), [pr#11181](#), Matt Benjamin)
- rgw: rgw\_file: fix rename cases and unify unlink ([pr#10271](#), Matt Benjamin)
- rgw: rgw\_file: fix set\_attrs operation ([pr#11159](#), Matt Benjamin)
- rgw: rgw\_file: refuse partial, out-of-order writes ([pr#10284](#), Matt Benjamin)
- rgw: rgw\_file: restore local definition of RGWLibFS gc interval ([pr#10756](#), Matt Benjamin)
- rgw: rgw\_file: unlock() must precede out label ([pr#10635](#), Matt Benjamin)
- rgw: right parenthesis is missing in radosgw-admin help message on caps ([pr#10947](#), Weibing Zhang)
- rgw: set correct instance on the object ([issue#17443](#), [pr#11270](#), Yehuda Sadeh)
- rgw: store oldest mdlog period in rados ([issue#16894](#), [pr#10558](#), Casey Bodley)

- rgw: test/multi.py add a destructive attr to tests ([pr#10401](#), Abhishek Lekshmanan)
- rgw: test/rgw: add -gateways-per-zone to test\_multi.py ([pr#10742](#), Casey Bodley)
- rgw: test\_multi.py avoid creating mds ([pr#10174](#), Abhishek Lekshmanan)
- rgw: test\_rgw\_bencode: null terminate strings before checking ([issue#16861](#), [pr#10510](#), Yehuda Sadeh)
- rgw: use endpoints from master zone instead of zonegroup ([issue#16834](#), [pr#10456](#), Casey Bodley)
- rgw: use the standard usage of string.find ([pr#10226](#), Yan Jun)
- rgw: verified: f23, subset of s3tests ([pr#10448](#), Pritha Srivastava)
- rgw: verified ([pr#10000](#), weiqiaomiao)
- rgw: verified non-regression (MS AD) ([pr#10597](#), Pritha Srivastava)
- rgw: verified: autobuild ([issue#16928](#), [pr#10579](#), Robin H. Johnson)
- rgw: verified: MS AD ([pr#10307](#), Pritha Srivastava)
- rgw: verified: f23 ([pr#10882](#), Michal Jarzabek)
- rgw: verified: f23 ([pr#10858](#), Weibing Zhang)
- rgw: verified: f23 ([pr#10822](#), Yan Jun)
- rgw: verified: f23 ([pr#10929](#), Weibing Zhang)
- rgw: wip: rgw multisite: preserve zone's extra pool ([issue#16712](#), [pr#10397](#), Abhishek Lekshmanan)
- rgw: work around curl\_multi\_wait bug with non-blocking reads ([issue#15915](#), [issue#16695](#), [pr#10998](#), Casey Bodley)
- rgw: add a s3 API of make torrent for a object ([pr#10396](#), zhouruisong)
- rgw: add a s3 API of make torrent for a object ([pr#9589](#), zhouruisong)
- rgw:bucket check remove \_multipart\_ prefix ([pr#6501](#), Weijun Duan)
- rgw:clean unuse bufferlist ([pr#10232](#), weiqiaomiao)
- rgw:fix rgw boot failed after upgrade to master latest version ([pr#10409](#), weiqiaomiao)
- rgw:lifecycle feature [rebased] ([pr#9737](#), Ji Chen, Daniel Gryniewicz)
- rgw: rgw/rgw\_rados.h: remove unneeded class C\_Tick ([pr#10954](#), Michal Jarzabek)

- rgw: ext\_mime\_map\_init add string describing for error number ([pr#9807](#), Yan Jun)
- tests: Add test for global static non-POD segfault ([pr#10486](#), Brad Hubbard)
- tests: populate /dev/disk/by-partuuid for scsi\_debug ([issue#17100](#), [pr#10824](#), Loic Dachary)
- tests: use a fixture for memstore clone testing ([pr#11190](#), Kefu Chai)
- tests: run-\*make-check.sh: Make DRY\_RUN actually mean a dry run ([pr#11074](#), Brad Hubbard)
- tests: run-cmake-check.sh: Actually run the tests ([pr#11075](#), Brad Hubbard)
- tests: run-cmake-check.sh: Init submodules ([pr#11091](#), Brad Hubbard)
- tests: run-make-check.sh: Make DRY\_RUN actually do a dry run ([pr#11092](#), Brad Hubbard)
- tests: run-make-check.sh: pass args to do\_cmake.sh ([pr#10701](#), John Coyle)
- tests: unittest\_chain\_xattr: account for existing xattrs ([issue#16025](#), [pr#11109](#), Dan Mick)
- tests: src/test/cli/\* tests: POSIX Convert grep -P to grep -E ([pr#10319](#), Willem Jan Withagen)
- test: ceph\_test\_msgr: fix circular locking dependency ([issue#16955](#), [pr#10612](#), Kefu Chai)
- test: cli/crushtool: fix the test of compile-decompile-recompile.t ([issue#17306](#), [pr#11173](#), Kefu Chai)
- test: libcephfs: fix gcc sys/fcntl.h warnings ([pr#10126](#), John Coyle)
- test: librados: rados\_connect() should succeed ([issue#17087](#), [pr#10806](#), Kefu Chai)
- test: mds: add fs dump in test\_ceph\_argparse.py ([pr#10347](#), huanwen ren)
- test: simple\_dispatcher.cc: remove unused variable ([pr#9932](#), Michal Jarzabek)
- test: store\_test: tidy-up SyntheticWorkloadState class ([pr#10775](#), xie xingguo)
- test: More portable use of mmap(MAP\_ANON) ([pr#10557](#), Willem Jan Withagen)
- test: Removeall merged after print\_function commit needs a fix ([pr#10535](#), David Zafman)
- test: ceph-disk.sh do not kill all daemons ([issue#16729](#), [pr#10346](#), Kefu Chai)
- test: cephtool/test.sh: fix expect\_false() calls ([pr#10133](#), Kefu Chai)
- test: fix usage info of omapbench ([pr#10089](#), Wanlong Gao)

- test: remove ceph\_test\_rados\_api\_tmap\_migrate ([issue#16144](#), [pr#10256](#), Kefu Chai)
- test: test\_{compression\_plugin,async\_compressor}: do not copy plugins ([pr#10153](#), Kefu Chai)
- test: test\_rados\_tool.sh: Make script work under ctest ([pr#10166](#), Willem Jan Withagen)
- test: qa/workunits/cephtool/test.sh: fix omission of ceph-command ([pr#10979](#), Willem Jan Withagen)
- test: qa/workunits/cephtool/test.sh: s/TMPDIR/TEMP\_DIR/ ([pr#10306](#), Kefu Chai)
- test: qa/workunits/cephtool/test.sh: use absolute path for TEMP\_DIR ([pr#10430](#), Kefu Chai)
- tools: New “removeall” used to remove head with snapshots ([pr#10098](#), David Zafman)
- tools: do not closed stdout ; fix overload of “<” operator ([pr#9290](#), xie xingguo)
- tools: fix the core dump when get the crushmap do not exist ([pr#10451](#), song baisen)
- tools: rebuild monstore ([issue#17179](#), [pr#10933](#), Kefu Chai)
- tools: use TextTable for “rados df” plain output ([pr#9362](#), xie xingguo)
- tools: fio engine for objectstore ([pr#10267](#), Casey Bodley, Igor Fedotov, Daniel Gollub)
- tools: rados/client: fix typo ([pr#10493](#), Yan Jun)
- tools: rados/client: fix waiting on the condition variable more efficient. ([pr#9939](#), Yan Jun)
- tools: tools/rebuild\_mondb: kill comipling warning and other fixes ([pr#11117](#), xie xingguo)
- tools: authtool: Enhance argument combinations validation ([issue#2904](#), [pr#9704](#), Brad Hubbard)
- tools: ceph-disk: change ownership of initfile to ceph:ceph ([issue#16280](#), [pr#9688](#), Shylesh Kumar)
- test: ceph\_test\_rados\_api\_tmap\_migrate: remove test for tmap\_upgrade ([pr#10234](#), Kefu Chai)

## v10.2.11 Jewel

This point releases brings a number of important bugfixes and has a few important security fixes. This is expected to be the last Jewel release. We recommend all Jewel 10.2.x users to upgrade.

## Notable Changes

- CVE 2018-1128: auth: cephx authorizer subject to replay attack ([issue#24836](#), Sage Weil)
- CVE 2018-1129: auth: cephx signature check is weak ([issue#24837](#), Sage Weil)
- CVE 2018-10861: mon: auth checks not correct for pool ops ([issue#24838](#), Jason Dillaman)
- The RBD C API's `rbd_discard` method and the C++ API's `Image::discard` method now enforce a maximum length of 2GB. This restriction prevents overflow of the result code.
- New OSDs will now use rocksdb for omap data by default, rather than leveldb. omap is used by RGW bucket indexes and CephFS directories, and when a single leveldb grows to 10s of GB with a high write or delete workload, it can lead to high latency when leveldb's single-threaded compaction cannot keep up. rocksdb supports multiple threads for compaction, which avoids this problem.
- The CephFS client now catches failures to clear dentries during startup and refuses to start as consistency and untrimmable cache issues may develop. The new option `client_die_on_failed_dentry_invalidate` (default: true) may be turned off to allow the client to proceed (dangerous!).
- In 10.2.10 and earlier releases, keyring caps were not checked for validity, so the caps string could be anything. As of 10.2.11, caps strings are validated and providing a keyring with an invalid caps string to, e.g., "ceph auth add" will result in an error.

## Changelog

- admin: bump sphinx to 1.6 ([issue#21717](#), [pr#18166](#), Kefu Chai, Alfredo Deza)
- auth: ceph auth add does not sanity-check caps ([issue#22525](#), [pr#21367](#), Jing Li, Nathan Cutler, Kefu Chai, Sage Weil)
- build/ops: rpm: bump epoch ahead of ceph-common in RHEL base ([issue#20508](#), [pr#21190](#), Ken Dreyer)

- build/ops: upstart: radosgw-all does not start on boot if ceph-base is not installed ([issue#18313](#), [pr#16294](#), Ken Dreyer)
- ceph\_authtool: add mode option ([issue#23513](#), [pr#21197](#), Sébastien Han)
- ceph-disk: factor out the retry logic into a decorator ([issue#21728](#), [pr#18169](#), Kefu Chai)
- ceph-disk: fix -runtime omission when enabling `ceph-osd@$ID.service` units for device-backed OSDs ([issue#21498](#), [pr#17942](#), Carl Xiong)
- ceph-disk flake8 test fails on very old, and very new, versions of flake8 ([issue#22207](#), [pr#19153](#), Nathan Cutler)
- cephfs: ceph.in: pass RADOS inst to LibCephFS ([issue#21406](#), [issue#21967](#), [pr#19907](#), Patrick Donnelly)
- cephfs: client::mkdirs not handle well when two clients send mkdir request for a same dir ([issue#20592](#), [pr#20271](#), dongdong tao)
- cephfs: client: prevent fallback to remount when dentry\_invalidate\_cb is true but root->dir is NULL ([issue#23211](#), [pr#21189](#), Zhi Zhang)
- cephfs: fix tmap\_upgrade crash ([issue#23529](#), [pr#21208](#), "Yan, Zheng")
- cephfs: fuse client: ::rmdir() uses a deleted memory structure of dentry leads ... ([issue#22536](#), [pr#19993](#), YunfeiGuan)
- cephfs-journal-tool: add "set pool\_id" option ([issue#22631](#), [pr#20111](#), dongdong tao)
- cephfs-journal-tool: move shutdown to the deconstructor of MDSUtility ([issue#22734](#), [pr#20333](#), dongdong tao)
- cephfs: osdc: "FAILED assert(bh->last\_write\_tid > tid)" in powercycle-wip-yuri-master-1.19.18-distro-basic-smithi ([issue#22741](#), [pr#20312](#), "Yan, Zheng")
- cephfs: osdc/Journaler: make sure flush() writes enough data ([issue#22824](#), [pr#20435](#), "Yan, Zheng")
- cephfs: Processes stuck waiting for write with ceph-fuse ([issue#22008](#), [issue#22207](#), [pr#19141](#), "Yan, Zheng")
- ceph-fuse: failure to remount in startup test does not handle `client_die_on_failed_remount` properly ([issue#22269](#), [pr#21162](#), Patrick Donnelly)
- ceph.in: bypass codec when writing raw binary data ([issue#23185](#), [pr#20763](#), Oleh Prypin)
- ceph-objectstore-tool command to trim the pg log ([issue#23242](#), [pr#20882](#), Josh Durgin, David Zafman)

- ceph-objectstore-tool: “\$OBJ get-omaphdr” and “\$OBJ list-omap” scan all pgs instead of using specific pg ([issue#21327](#), [pr#20284](#), David Zafman)
- ceph.restart + ceph\_manager.wait\_for\_clean is racy ([issue#15778](#), [pr#20508](#), Warren Usui, Sage Weil)
- ceph\_volume\_client: fix setting caps for IDs ([issue#21501](#), [pr#18084](#), Ramana Raja)
- class rbd.Image discard--OSError: [errno 2147483648] error discarding region ([issue#16465](#), [issue#21966](#), [pr#20287](#), Nathan Cutler, Huan Zhang, Jason Dillaman)
- cli/crushtools/build.t sometimes fails in jenkins' make check run ([issue#21758](#), [pr#21158](#), Kefu Chai)
- client reconnect gather race ([issue#22263](#), [pr#21163](#), “Yan, Zheng”)
- client: release revoking Fc after invalidate cache ([issue#22652](#), [pr#19975](#), “Yan, Zheng”)
- client: set client\_try\_dentry\_invalidate to false by default ([issue#21423](#), [pr#17925](#), “Yan, Zheng”)
- [cli] rename of non-existent image results in seg fault ([issue#21248](#), [pr#20280](#), Jason Dillaman)
- CLI unit formatting tests are broken ([issue#24733](#), [pr#22913](#), Jason Dillaman)
- common: compute SimpleLRU's size with contents.size() instead of lru.... ([issue#22613](#), [pr#19978](#), Xuehan Xu)
- common/config: set rocksdb\_cache\_size to OPT\_U64 ([issue#22104](#), [pr#18850](#), Vikhyat Umrao, liuhongtong)
- common: fix typo in rados bench write JSON output ([issue#24199](#), [pr#22407](#), Sandor Zeestraten)
- config: lower default omap entries recovered at once ([issue#21897](#), [pr#19927](#), Josh Durgin)
- core: Addition of online osd ‘omap’ compaction command ([issue#19592](#), [pr#17101](#), liuchang0812, Sage Weil)
- core: global/signal\_handler.cc: fix typo ([issue#21432](#), [pr#17883](#), Kefu Chai)
- core: librados: Double free in rados\_getxattrs\_next ([issue#22042](#), [pr#20381](#), Gu Zhongyan)
- core: Objecter::C\_ObjectOperation\_sparse\_read throws/catches exceptions on - ENOENT ([issue#21844](#), [pr#18743](#), Jason Dillaman)
- Deleting a pool with active notify linger ops can result in seg fault ([issue#23966](#), [pr#22188](#), Kefu Chai, Jason Dillaman)

- doc: clarify Path Restriction instructions ([issue#16906](#), [pr#19795](#), huanwen ren)
- doc: clarify Path Restriction instructions ([issue#16906](#), [pr#19840](#), Drunkard Zhang)
- doc: remove region from INSTALL CEPH OBJECT GATEWAY ([issue#21610](#), [pr#18303](#), Orit Wasserman)
- Filestore rocksdb compaction readahead option not set by default ([issue#21505](#), [pr#20446](#), Mark Nelson)
- follow-on: osd: be\_select\_auth\_object() sanity check oi soid ([issue#20471](#), [pr#20622](#), David Zafman)
- HashIndex: randomize split threshold by a configurable amount ([issue#15835](#), [pr#19906](#), Josh Durgin)
- include/fs\_types: fix unsigned integer overflow ([issue#22494](#), [pr#19611](#), runsis)
- install-deps.sh: point gcc to the one shipped by distro ([issue#22220](#), [pr#19461](#), Kefu Chai)
- install-deps.sh: readlink /usr/bin/gcc not /usr/bin/x86\_64-linux-gnu-gcc ([issue#22220](#), [pr#19521](#), Kefu Chai)
- install-deps.sh: update g++ symlink also ([issue#22220](#), [pr#19656](#), Kefu Chai)
- journal: Message too long error when appending journal ([issue#23526](#), [pr#21215](#), Mykola Golub)
- [journal] tags are not being expired if no other clients are registered ([issue#21960](#), [pr#20282](#), Jason Dillaman)
- legal: remove doc license ambiguity ([issue#23336](#), [pr#20999](#), Nathan Cutler)
- librados: copy out data to users' buffer for xio ([issue#20616](#), [pr#17594](#), Vu Pham)
- librbd: cannot clone all image-metas if we have more than 64 key/value pairs ([issue#21814](#), [pr#21228](#), PCzhangPC)
- librbd: cannot copy all image-metas if we have more than 64 key/value pairs ([issue#21815](#), [pr#21203](#), PCzhangPC)
- librbd: create+truncate for whole-object layered discards ([issue#23285](#), [pr#21219](#), Jason Dillaman)
- librbd: list\_children should not attempt to refresh image ([issue#21670](#), [pr#21224](#), Jason Dillaman)
- librbd: object map batch update might cause OSD suicide timeout ([issue#22716](#), [issue#21797](#), [pr#21220](#), Song Shun, Jason Dillaman)

- librbd: set deleted parent pointer to null ([issue#22158](#), [pr#19098](#), Jason Dillaman)
- log: Fix AddressSanitizer: new-delete-type-mismatch ([issue#23324](#), [pr#21084](#), Brad Hubbard)
- mds: FAILED assert(get\_version() < pv) in CDir::mark\_dirty ([issue#21584](#), [pr#21156](#), Yan, Zheng, "Yan, Zheng")
- mds: fix dump last\_sent ([issue#22562](#), [pr#19961](#), dongdong tao)
- mds: fix integer overflow ([issue#21067](#), [pr#17188](#), Henry Chang)
- mds: fix scrub crash ([issue#22730](#), [pr#20335](#), dongdong tao)
- mds: session reference leak ([issue#22821](#), [pr#21175](#), Nathan Cutler, "Yan, Zheng")
- mds: unbalanced auth\_pin/auth\_unpin in RecoveryQueue code ([issue#22647](#), [pr#20067](#), "Yan, Zheng")
- mds: underwater dentry check in CDir::\_omap\_fetched is racy ([issue#23032](#), [pr#21185](#), Yan, Zheng)
- mon/LogMonitor: call no\_reply() on ignored log message ([issue#24180](#), [pr#22431](#), Sage Weil)
- mon/MDSMonitor: no\_reply on MMDSLoadTargets ([issue#23769](#), [pr#22189](#), Sage Weil)
- mon/OSDMonitor.cc: fix expected\_num\_objects interpret error ([issue#22530](#), [pr#22050](#), Yang Honggang)
- mon/OSDMonitor: fix dividing by zero in OSDUtilizationDumper ([issue#22662](#), [pr#20344](#), Mingxin Liu)
- ObjectStore/StoreTest.FiemapHoles/3 fails with kstore ([issue#21716](#), [pr#20143](#), Kefu Chai, Ning Yao)
- osd: also check the exsistence of clone obc for "CEPH\_SNAPDIR" requests ([issue#17445](#), [pr#17707](#), Xuehan Xu)
- osdc/Objecter: prevent double-invocation of linger op callback ([issue#23872](#), [pr#21754](#), Jason Dillaman)
- osd: objecter sends out of sync with pg epochs for proxied ops ([issue#22123](#), [pr#20518](#), Sage Weil)
- osd ops (sent and?) arrive at osd out of order ([issue#19133](#), [issue#19139](#), [pr#17893](#), Jianpeng Ma, Sage Weil)
- osd: OSDMap cache assert on shutdown ([issue#21737](#), [pr#21184](#), Greg Farnum)
- osd: osd\_scrub\_during\_recovery only considers primary, not replicas ([issue#18206](#),

[pr#17815](#), David Zafman)

- osd/PrimaryLogPG: dump snap\_trimq size ([issue#22448](#), [pr#21200](#), Piotr Dałek)
- osd: recover\_replicas: object added to missing set for backfill, but is not in recovering, error! ([issue#18162](#), [issue#14513](#), [pr#18690](#), huangjun, Adam C. Emerson, David Zafman)
- osd: replica read can trigger cache promotion ([issue#20919](#), [pr#21199](#), Sage Weil)
- osd: update heartbeat peers when a new OSD is added ([issue#18004](#), [pr#20108](#), Pan Liu)
- performance: Only scan for omap corruption once ([issue#21328](#), [pr#18951](#), David Zafman)
- qa: failures from pjd fstest ([issue#21383](#), [pr#21152](#), "Yan, Zheng")
- qa: src/test/libcephfs/test.cc:376: Expected: (len) > (0), actual: -34 vs 0 ([issue#22221](#), [pr#21172](#), Patrick Donnelly)
- qa: use xfs instead of btrfs w/ filestore ([issue#20169](#), [issue#20911](#), [pr#18165](#), Sage Weil)
- qa: use xfs instead of btrfs w/ filestore ([issue#21481](#), [pr#17847](#), Patrick Donnelly)
- radosgw: fix awsv4 header line sort order ([issue#21607](#), [pr#18080](#), Marcus Watts)
- rbd: clean up warnings when mirror commands used on non-setup pool ([issue#21319](#), [pr#21227](#), Jason Dillaman)
- rbd: disk usage on empty pool no longer returns an error message ([issue#22200](#), [pr#19186](#), Jason Dillaman)
- [rbd] image-meta list does not return all entries ([issue#21179](#), [pr#20281](#), Jason Dillaman)
- rbd: is\_qemu\_running in qemu\_rebuild\_object\_map.sh and qemu\_dynamic\_features.sh may return false positive ([issue#23502](#), [pr#21207](#), Mykola Golub)
- rbd: [journal] allocating a new tag after acquiring the lock should use on-disk committed position ([issue#22945](#), [pr#21206](#), Jason Dillaman)
- rbd: librbd: filter out potential race with image rename ([issue#18435](#), [pr#19855](#), Jason Dillaman)
- rbd ls -l crashes with SIGABRT ([issue#21558](#), [pr#19801](#), Jason Dillaman)
- rbd-mirror: cluster watcher should ensure it has latest OSD map ([issue#22461](#), [pr#19644](#), Jason Dillaman)

- rbd-mirror: fix potential infinite loop when formatting status message ([issue#22932](#), [pr#20418](#), Mykola Golub)
- rbd-mirror: ignore permission errors on rbd\_mirroring object ([issue#20571](#), [pr#21225](#), Jason Dillaman)
- rbd-mirror: strip environment/CLI overrides for remote cluster ([issue#21894](#), [pr#21223](#), Jason Dillaman)
- [rbd-nbd] Fedora does not register resize events ([issue#22131](#), [pr#19115](#), Jason Dillaman)
- rbd-nbd: fix ebusy when do map ([issue#23528](#), [pr#21232](#), Li Wang)
- rbd: possible deadlock in various maintenance operations ([issue#22120](#), [pr#20285](#), Jason Dillaman)
- rbd: rbd crashes during map ([issue#21808](#), [pr#18843](#), Peter Keresztes Schmidt)
- rbd: rbd-mirror split brain test case can have a false-positive failure until teuthology ([issue#22485](#), [pr#21205](#), Jason Dillaman)
- rbd: TestLibRBD.RenameViaLockOwner may still fail with -ENOENT ([issue#23068](#), [pr#20627](#), Mykola Golub)
- repair\_test fails due to race with osd start ([issue#20705](#), [pr#20146](#), Sage Weil)
- rgw: 15912 15673 (Fix duplicate tag removal during GC, cls/refcount: store and use list of retired tags) ([issue#20107](#), [pr#16708](#), Jens Rosenboom)
- rgw: abort in listing mapped nbd devices when running in a container ([issue#22012](#), [issue#22011](#), [pr#20286](#), Li Wang, Pan Liu)
- rgw: add ability to sync user stats from admin api ([issue#21301](#), [pr#20179](#), Nathan Johnson)
- rgw: add cors header rule check in cors option request ([issue#22002](#), [pr#19057](#), yuliyang)
- rgw: add radosgw-admin sync error trim to trim sync error log ([issue#23287](#), [pr#21210](#), fang yuxiang)
- rgw: add xml output header in RGWCopyObj\_ObjStore\_S3 response msg ([issue#22416](#), [pr#19887](#), Enming Zhang)
- rgw: automated trimming of datalog and mdlog ([issue#18227](#), [pr#20061](#), Casey Bodley)
- rgw: bi list entry count incremented on error, distorting error code ([issue#21205](#), [pr#18207](#), Nathan Cutler)
- rgw: boto3 v4 SignatureDoesNotMatch failure due to sorting of sse-kms headers

([issue#21832](#), [pr#18772](#), Nathan Cutler)

- rgw: bucket resharding should not update bucket ACL or user stats ([issue#22124](#), [pr#20421](#), Orit Wasserman)
- rgw: copying part without http header x-amz-copy-source-range will be mistaken for copying object ([issue#22729](#), [pr#21294](#), Malcolm Lee)
- rgw: core dump, recursive lock of RGWKeystoneTokenCache ([issue#23171](#), [pr#20639](#), Mark Kogan, Adam Kupczyk)
- rgw: data sync of versioned objects, note updating bi marker ([issue#18885](#), [pr#21213](#), Yehuda Sadeh)
- rgw: dont log EBUSY errors in 'sync error list' ([issue#22473](#), [pr#19908](#), Casey Bodley)
- rgw: ECANCELED in rgw\_get\_system\_obj() leads to infinite loop ([issue#17996](#), [pr#20561](#), Yehuda Sadeh)
- rgw: file deadlock on lru evicting ([issue#22736](#), [pr#20076](#), Matt Benjamin)
- rgw: file write error ([issue#21455](#), [pr#18304](#), Yao Zongyou)
- rgw: fix chained cache invalidation to prevent cache size growth ([issue#22410](#), [pr#19469](#), Mark Kogan)
- rgw: fix doubled underscore with s3/swift server-side copy ([issue#22529](#), [pr#19747](#), Matt Benjamin)
- rgw: fix GET website response error code ([issue#22272](#), [pr#19488](#), Dmitry Plyakin)
- rgw: fix index update in dir\_suggest\_changes ([issue#24280](#), [pr#22677](#), Tianshan Qu)
- rgw: fix marker encoding problem ([issue#20463](#), [pr#17731](#), Orit Wasserman, Marcus Watts)
- rgw: fix swift anonymous access ([issue#22259](#), [pr#19194](#), Marcus Watts)
- rgw: Fix swift object expiry not deleting objects ([issue#22084](#), [pr#18925](#), Pavan Rallabhandi)
- rgw: fix the bug that part's index can't be removed after completing ([issue#19604](#), [pr#16763](#), Zhang Shaowen, Matt Benjamin)
- rgw: fix the max-uploads parameter not work ([issue#22825](#), [pr#20479](#), Xin Liao)
- rgw: inefficient buffer usage for PUTs ([issue#23207](#), [pr#21098](#), Marcus Watts)
- rgw: libcurl & ssl fixes ([issue#22951](#), [issue#23203](#), [issue#23162](#), [pr#20749](#), Marcus Watts, Abhishek Lekshmanan, Jesse Williamson)

- rgw: list bucket which enable versioning get wrong result when user marker ([issue#21500](#), [pr#20291](#), yuliyang)
- rgw: log includes zero byte sometimes ([issue#20037](#), [pr#17151](#), Abhishek Lekshmanan)
- rgw: make init env methods return an error ([issue#23039](#), [pr#20800](#), Abhishek Lekshmanan)
- RGW: Multipart upload may double the quota ([issue#21586](#), [pr#18121](#), Sibei Gao, Matt Benjamin)
- rgw: multisite: data sync status advances despite failure in RGWListBucketIndexesCR ([issue#21735](#), [pr#20269](#), Casey Bodley)
- rgw: multisite: Get bucket location which is located in another zonegroup, will return 301 Moved Permanently ([issue#21125](#), [pr#18305](#), Shasha Lu, lvshuhua, Jiaying Ren)
- rgw: null instance mtime incorrect when enable versioning ([issue#21743](#), [pr#20262](#), Shasha Lu)
- rgw: radosgw-admin: add an option to reset user stats ([issue#23335](#), [issue#23322](#), [pr#20877](#), Abhishek Lekshmanan)
- rgw: release cls lock if taken in RGWCompleteMultipart ([issue#21596](#), [issue#22368](#), [pr#18116](#), Casey Bodley, Matt Benjamin)
- rgw: resharding needs to set back the bucket ACL after link ([issue#22742](#), [pr#20039](#), Orit Wasserman)
- rgw: resolve Random 500 errors in Swift PutObject (22517) ([issue#22517](#), [issue#21560](#), [pr#19769](#), Adam C. Emerson, Matt Benjamin)
- rgw: rgw\_file: recursive lane lock can occur in LRU drain ([issue#20374](#), [pr#17149](#), Matt Benjamin)
- rgw: S3 POST policy should not require Content-Type ([issue#20201](#), [pr#19635](#), Matt Benjamin)
- rgw: s3website error handler uses original object name ([issue#23201](#), [issue#20307](#), [pr#21100](#), liuhong, Casey Bodley)
- rgw: segfaults after running radosgw-admin data sync init ([issue#22083](#), [pr#19783](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: segmentation fault when starting radosgw after reverting .rgw.root ([issue#21996](#), [pr#20292](#), Orit Wasserman, Casey Bodley)
- rgw: stale bucket index entry remains after object deletion ([issue#22555](#), [pr#20293](#), J. Eric Ivancich)

- rgw: system user can't delete bucket completely ([issue#22248](#), [pr#21212](#), Casey Bodley)
- rgw: tcmalloc ([issue#23469](#), [pr#21073](#), Matt Benjamin)
- rgw: update the max-buckets when the quota is uploaded ([issue#22745](#), [pr#20496](#), zhaokun)
- rgw: user creation can overwrite existing user even if different uid is given ([issue#21685](#), [pr#20074](#), Casey Bodley)
- RHEL 7.3 Selinux denials at OSD start ([issue#19200](#), [pr#18780](#), Boris Ranto)
- scrub errors not cleared on replicas can cause inconsistent pg state when replica takes over primary ([issue#23267](#), [pr#21194](#), David Zafman)
- snapset xattr corruption propagated from primary to other shards ([issue#20186](#), [issue#18409](#), [issue#21907](#), [pr#20331](#), David Zafman)
- systemd: Add explicit Before=ceph.target ([issue#21477](#), [pr#17841](#), Tim Serong)
- table of contents doesn't render for luminous/jewel docs ([issue#23780](#), [pr#21503](#), Alfredo Deza)
- test: Adjust for Jewel quirk caused of differences with master ([issue#23006](#), [pr#20463](#), David Zafman)
- test/CMakeLists: disable test\_pidfile.sh ([issue#20975](#), [pr#20557](#), Sage Weil)
- test\_health\_warnings.sh can fail ([issue#21121](#), [pr#20289](#), Sage Weil)
- test/librbd: fixed metadata tests under upgrade scenarios ([issue#21911](#), [pr#18548](#), Jason Dillaman)
- test/librbd: utilize unique pool for cache tier testing ([issue#11502](#), [pr#20524](#), Jason Dillaman)
- tests: rbd\_mirror\_helpers.sh request\_resync\_image function saves image id to wrong variable ([issue#21663](#), [pr#19804](#), Jason Dillaman)
- tests: test\_admin\_socket.sh may fail on wait\_for\_clean ([issue#23499](#), [pr#21125](#), Mykola Golub)
- tests: tests/librbd: updated test\_notify to handle new release lock semantics ([issue#21912](#), [pr#18560](#), Jason Dillaman)
- tests: unittest\_pglog timeout ([issue#23504](#), [issue#18030](#), [pr#21135](#), Nathan Cutler, Loic Dachary)
- tools: ceph-objectstore-tool set-size should clear data-digest ([issue#22112](#), [pr#20070](#), David Zafman)

- Ubuntu amd64 client can not discover the ubuntu arm64 ceph cluster ([issue#19705](#), [pr#18294](#), Kefu Chai)

## v10.2.10 Jewel

---

This point release brings a number of important bugfixes in all major components of Ceph, we recommend all Jewel 10.2.x users to upgrade.

For a detailed list of changes refer to :download: the complete changelog  
<./changelog/v10.2.10.txt>

## Notable Changes

---

- build/ops: Add fix subcommand to ceph-disk, fix SELinux denials, and speed up upgrade from non-SELinux enabled ceph to an SELinux enabled one ([issue#20077](#), [issue#20184](#), [issue#19545](#), [pr#14346](#), Boris Ranto)
- build/ops: deb: Fix logrotate packaging ([issue#19938](#), [pr#15428](#), Nathan Cutler)
- build/ops: extended, customizable systemd ceph-disk timeout ([issue#18740](#), [pr#15051](#), Alexey Sheplyakov)
- build/ops: rpm: fix python-Sphinx package name for SUSE ([issue#19924](#), [pr#15196](#), Nathan Cutler, Jan Matejek)
- build/ops: rpm: set subman cron attributes in spec file ([issue#20074](#), [pr#15473](#), Thomas Serlin)
- cephfs: ceph-fuse segfaults at mount time, assert in ceph::log::Log::stop ([issue#18157](#), [pr#16963](#), Greg Farnum)
- cephfs: df reports negative disk “used” value when quota exceed ([issue#20178](#), [pr#16151](#), John Spray)
- cephfs: get\_quota\_root sends lookupname op for every buffered write ([issue#20945](#), [pr#17396](#), Dan van der Ster)
- cephfs: osdc/Filer: truncate large file party by party ([issue#19755](#), [pr#15442](#), “Yan, Zheng”)
- core: an OSD was seen getting ENOSPC even with osdfailsafe\_full\_ratio passed ([issue#20544](#), [issue#16878](#), [issue#19733](#), [issue#15912](#), [pr#15050](#), Sage Weil, David Zafman)
- core: disable skewed utilization warning by default ([issue#20730](#), [pr#17210](#), David Zafman)
- core: interval\_set: optimize intersect\_of insert operations ([issue#21229](#),

pr#17514, Zac Medico)

- core: kv: let ceph\_logger destructed after db reset ([issue#21336](#), [pr#17626](#), wumingqiao)
- core: test\_envlibrados\_for\_rocksdb.yaml fails on crypto restart ([issue#19741](#), [pr#16293](#), Kefu Chai)
- libradosstriper silently fails to delete empty objects in jewel ([issue#20325](#), [pr#15760](#), Stan K)
- librbd: fail IO request when exclusive lock cannot be obtained ([issue#20168](#), [issue#21251](#), [pr#17402](#), Jason Dillaman)
- librbd: prevent self-blacklisting during break lock ([issue#18666](#), [pr#17412](#), Jason Dillaman)
- librbd: reacquire lock should update lock owner client id ([issue#19929](#), [pr#17385](#), Jason Dillaman)
- mds: damage reporting by ino number is useless ([issue#18509](#), [issue#16016](#), [pr#14699](#), John Spray, Michal Jarzabek)
- mds: log rotation doesn't work if mds has respawned ([issue#19291](#), [pr#14673](#), Patrick Donnelly)
- mds: save projected path into inode\_t::stray\_prior\_path ([issue#20340](#), [pr#16150](#), "Yan, Zheng")
- mon: crash on shutdown, lease\_ack\_timeout event ([issue#19825](#), [pr#15083](#), Kefu Chai, Michal Jarzabek, Alexey Sheplyakov)
- mon: Disallow enabling 'hashpspool' option to a pool without some kind of -i-understand-this-will-remap-all-pgs flag ([issue#18468](#), [pr#13507](#), Vikhyat Umrao)
- mon: factor mon\_osd\_full\_ratio into MAX AVAIL calc ([issue#18522](#), [pr#15236](#), Sage Weil)
- mon: fail to form large quorum; msg/async busy loop ([issue#20230](#), [pr#15726](#), Haomai Wang, Michal Jarzabek)
- mon: fix force\_pg\_create pg stuck in creating bug ([issue#18298](#), [pr#17008](#), Alexey Sheplyakov)
- mon: osd crush set crushmap need sanity check ([issue#19302](#), [pr#16144](#), Loic Dachary)
- osd: Add heartbeat message for Jumbo Frames (MTU 9000) ([issue#20087](#), [issue#20323](#), [pr#16059](#), Piotr Dałek, Sage Weil, Greg Farnum)
- osd: fix infinite loops in fiemap ([issue#19996](#), [pr#15189](#), Sage Weil, Ning Yao)

- osd: leaked MOSDMap ([issue#18293](#), [pr#14943](#), Sage Weil)
- osd: objecter full\_try behavior not consistent with osd ([issue#19430](#), [pr#15474](#), Sage Weil)
- osd: omap threadpool heartbeat is only reset every 100 values ([issue#20375](#), [pr#16167](#), Josh Durgin)
- osd: osd\_internal\_types: wake snaptrimmer on put\_read lock, too ([issue#19131](#), [pr#16015](#), Sage Weil)
- osd: PrimaryLogPG: do not call on\_shutdown() if (pg.deleting) ([issue#19902](#), [pr#15065](#), Kefu Chai)
- osd: rados ls on pool with no access returns no error ([issue#20043](#), [issue#19790](#), [pr#16473](#), Nathan Cutler, Kefu Chai, John Spray, Sage Weil, Brad Hubbard)
- osd: ReplicatedPG: solve cache tier osd high memory consumption ([issue#20464](#), [pr#16169](#), Peng Xie)
- osd: Reset() snaptrimmer on shutdown and do not default-abort on leaked pg refs ([issue#19931](#), [pr#15322](#), Greg Farnum)
- osd: scrub\_to specifies clone ver, but transaction include head write ver ([issue#20041](#), [pr#16405](#), David Zafman)
- osd: unlock sdata\_op\_ordering\_lock with sdata\_lock hold to avoid missing wakeup signal ([issue#20427](#), [pr#15947](#), Alexey Sheplyakov)
- qa: add a sleep after restarting osd before "tell"ing it ([issue#16239](#), [pr#15475](#), Kefu Chai)
- rbd: api: is\_exclusive\_lock\_owner shouldn't return -EBUSY ([issue#20182](#), [pr#16296](#), Jason Dillaman)
- rbd: cli: ensure positional arguments exist before casting ([issue#20185](#), [pr#16295](#), Jason Dillaman)
- rbd: cli: map with cephx disabled results in error message ([issue#19035](#), [pr#16297](#), Jason Dillaman)
- rbd: default features should be negotiated with the OSD ([issue#17010](#), [pr#14874](#), Mykola Golub, Jason Dillaman)
- rbd: Enabling mirroring for a pool with clones may fail ([issue#19798](#), [issue#19130](#), [pr#14663](#), Mykola Golub, Jason Dillaman)
- rbd-mirror: image sync should send NOCACHE advise flag ([issue#17127](#), [pr#16285](#), Mykola Golub)
- rbd: object-map: batch updates during trim operation ([issue#17356](#), [pr#15460](#),

Mykola Golub, Venky Shankar, Nathan Cutler)

- rbd: Potential IO hang if image is flattened while read request is in-flight ([issue#19832](#), [pr#15464](#), Jason Dillaman)
- rbd: rbd\_clone\_copy\_on\_read ineffective with exclusive-lock ([issue#18888](#), [pr#16124](#), Nathan Cutler, Venky Shankar, Jason Dillaman)
- rbd: rbd-mirror: ensure missing images are re-synced when detected ([issue#19811](#), [pr#15488](#), Jason Dillaman)
- rbd: rbd-mirror: failover and fallback of unmodified image results in split-brain ([issue#19858](#), [pr#14977](#), Jason Dillaman)
- rbd: rbd-nbd: kernel reported invalid device size (0, expected 1073741824) ([issue#19871](#), [pr#15463](#), Mykola Golub)
- rgw: add the remove-x-delete feature to cancel swift object expiration ([issue#19074](#), [pr#14659](#), Jing Wenjun)
- rgw: aws4: add rgw\_s3\_auth\_aws4\_force\_boto2\_compat conf option ([issue#16463](#), [pr#17009](#), Javier M. Mellid)
- rgw: bucket index check in radosgw-admin removes valid index ([issue#18470](#), [pr#16856](#), Zhang Shaowen, Pavan Rallabhandi)
- rgw: cls: ceph::timespan tag\_timeout wrong units ([issue#20380](#), [pr#16289](#), Matt Benjamin)
- rgw: Custom data header support ([issue#19644](#), [pr#15966](#), Pavan Rallabhandi)
- rgw: datalog trim can't work as expected ([issue#20190](#), [pr#16299](#), Zhang Shaowen)
- rgw: Delete non-empty bucket in slave zonegroup ([issue#19313](#), [pr#15477](#), Zhang Shaowen)
- rgw: Do not decrement stats cache when the cache values are zero ([issue#20661](#), [issue#20934](#), [pr#16720](#), Aleksei Gutikov, Pavan Rallabhandi)
- rgw: fix crash caused by shard id out of range when listing data log ([issue#19732](#), [pr#15465](#), redickwang)
- rgw: fix hangs in RGWRealmReloader::reload on SIGHUP ([issue#20686](#), [pr#17281](#), fang.yuxiang)
- rgw: fix infinite loop in rest api for log list ([issue#20386](#), [pr#15988](#), xierui, Casey Bodley)
- rgw: fix race in RGWCompleteMultipart ([issue#20861](#), [pr#16767](#), Abhishek Varshney, Matt Benjamin)
- rgw: Fix up to 1000 entries at a time in check\_bad\_index\_multipart ([issue#20772](#),

pr#16880, Orit Wasserman, Matt Benjamin)

- rgw: folders starting with \_ underscore are not in bucket index ([issue#19432](#), [pr#16276](#), Giovani Rinaldi, Orit Wasserman)
- rgw: 'gc list -include-all' command infinite loop the first 1000 items ([issue#19978](#), [pr#15719](#), Shasha Lu, fang yuxiang)
- rgw: meta sync thread crash at RGWMetaSyncShardCR ([issue#20251](#), [pr#16711](#), fang yuxiang, Nathan Cutler)
- rgw: multipart copy-part remove '/' for s3 java sdk request header ([issue#20075](#), [pr#16266](#), donglingpeng)
- rgw: multipart parts on versioned bucket create versioned bucket index entries ([issue#19604](#), [issue#17964](#), [pr#17278](#), Zhang Shaowen)
- rgw: multisite: after CreateBucket is forwarded to master, local bucket may use different value for bucket index shards ([issue#19745](#), [pr#15450](#), Shasha Lu)
- rgw: multisite: bucket zonegroup redirect not working ([issue#19488](#), [pr#15448](#), Casey Bodley)
- rgw: multisite: fixes for meta sync across periods ([issue#18639](#), [pr#15556](#), Casey Bodley)
- rgw: multisite: lock is not released when RGWMetaSyncShardCR::full\_sync() fails to write marker ([issue#18077](#), [pr#17155](#), Zhang Shaowen)
- rgw: multisite: log\_meta on secondary zone causes continuous loop of metadata sync ([issue#20357](#), [issue#20244](#), [pr#17148](#), Orit Wasserman, Casey Bodley)
- rgw: multisite: memory leak on failed lease in RGWDataSyncShardCR ([issue#19861](#), [issue#19834](#), [issue#19446](#), [pr#15457](#), Casey Bodley, weiqiaomiao)
- rgw: multisite: operating bucket's acl&cors is not restricted on slave zone ([issue#16888](#), [pr#15453](#), Casey Bodley, Shasha Lu, Guo Zhandong)
- rgw: multisite: realm rename does not propagate to other clusters ([issue#19746](#), [pr#15454](#), Casey Bodley)
- rgw: multisite: rest api fails to decode large period on "period commit" ([issue#19505](#), [pr#15447](#), Casey Bodley)
- rgw: multisite: RGWPeriodPuller does not call RGWPeriod::reflect() on new period ([issue#19816](#), [issue#19817](#), [pr#17167](#), Casey Bodley)
- rgw: multisite: RGWRadosRemoveOmapKeysCR::request\_complete return val is wrong ([issue#20539](#), [pr#17156](#), Shasha Lu)
- rgw: not initialized pointer cause rgw crash with ec data pool ([issue#20542](#),

[pr#17164](#), Aleksei Gutikov, fang yuxiang)

- rgw: radosgw-admin: bucket rm with -bypass-gc and without -purge-data doesn't throw error message ([issue#20688](#), [pr#17159](#), Abhishek Varshney)
- rgw: radosgw-admin data sync run crash ([issue#20423](#), [pr#17165](#), Shasha Lu)
- rgw: radosgw-admin: fix bucket limit check argparse, div(0) ([issue#20966](#), [pr#16952](#), Matt Benjamin)
- rgw: reduce log level of 'storing entry at' in cls\_log ([issue#19835](#), [pr#15455](#), Willem Jan Withagen)
- rgw: remove unnecessary 'error in read\_id for object name: default' ([issue#19922](#), [pr#15197](#), weiqiaomiao)
- rgw: replace '+' with "%20" in canonical query string for s3 v4 auth ([issue#20501](#), [pr#16951](#), Zhang Shaowen, Matt Benjamin)
- rgw: rgw\_common.cc: modify the end check in RGWHTTPArgs::sys\_get ([issue#16072](#), [pr#16268](#), zhao kun)
- rgw: rgw\_file: cannot delete bucket w/uxattrs ([issue#20061](#), [issue#20047](#), [issue#19214](#), [issue#20045](#), [pr#15459](#), Matt Benjamin)
- rgw: rgw\_file: fix size and (c|m)time unix attrs in write\_finish ([issue#19653](#), [pr#15449](#), Matt Benjamin)
- rgw: rgw\_file: incorrect lane lock behavior in evict\_block() ([issue#21141](#), [pr#17597](#), Matt Benjamin)
- rgw: rgw\_file: prevent conflict of mkdir between restarts ([issue#20275](#), [pr#17147](#), Gui Hecheng)
- rgw: rgw\_file: v3 write timer does not close open handles ([issue#19932](#), [pr#15456](#), Matt Benjamin)
- rgw: Segmentation fault when exporting rgw bucket in nfs-ganesha ([issue#20663](#), [pr#17285](#), Matt Benjamin)
- rgw: send data-log list infinitely ([issue#20951](#), [pr#17287](#), fang.yuxiang)
- rgw: set latest object's acl failed ([issue#18649](#), [pr#15451](#), Zhang Shaowen)
- rgw: Truncated objects ([issue#20107](#), [pr#17166](#), Yehuda Sadeh)
- rgw: uninitialized memory is accessed during creation of bucket's metadata ([issue#20774](#), [pr#17280](#), Radoslaw Zarzynski)
- rgw: usage logging on tenanted buckets causes invalid memory reads ([issue#20779](#), [pr#17279](#), Radoslaw Zarzynski)

- rgw: user quota did not work well on multipart upload ([issue#19285](#), [issue#19602](#), [pr#17277](#), Zhang Shaowen)
- rgw: VersionIdMarker and NextVersionIdMarker are not returned when listing object versions ([issue#19886](#), [pr#16316](#), Zhang Shaowen)
- rgw: when uploading objects continuously into a versioned bucket, some objects will not sync ([issue#18208](#), [pr#15452](#), lvshuhua)
- tools: ceph cli: Rados object in state configuring race ([issue#16477](#), [pr#15762](#), Loic Dachary)
- tools: ceph-disk: dmcrypt cluster must default to ceph ([issue#20893](#), [pr#16870](#), Loic Dachary)
- tools: ceph-disk: don't activate suppressed journal devices ([issue#19489](#), [pr#16703](#), David Disseldorf)
- tools: ceph-disk: separate ceph-osd -check-needs-\* logs ([issue#19888](#), [pr#15503](#), Loic Dachary)
- tools: ceph-disk: systemd unit timesout too quickly ([issue#20229](#), [pr#17133](#), Loic Dachary)
- tools: ceph-disk: Use stdin for 'config-key put' command ([issue#21059](#), [pr#17084](#), Brad Hubbard, Loic Dachary, Sage Weil)
- tools: libradosstriper processes arbitrary printf placeholders in user input ([issue#20240](#), [pr#17574](#), Stan K)

## v10.2.9 Jewel

---

This point release fixes a regression introduced in v10.2.8.

We recommend that all Jewel users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- cephfs: Damaged MDS with 10.2.8 ([issue#20599](#), [pr#16282](#), Nathan Cutler)

## v10.2.8 Jewel

---

This point release brought a number of important bugfixes in all major components of Ceph. However, it also introduced a regression that could cause MDS damage, and a new release, v10.2.9, was published to address this. Therefore, Jewel users should *not*

upgrade to this version - instead, we recommend upgrading directly to v10.2.9.

For more detailed information, see [the complete changelog](#).

## OSD Removal Caveat

There was a bug introduced in Jewel (#19119) that broke the mapping behavior when an “out” OSD that still existed in the CRUSH map was removed with ‘osd rm’. This could result in ‘misdirected op’ and other errors. The bug is now fixed, but the fix itself introduces the same risk because the behavior may vary between clients and OSDs. To avoid problems, please ensure that all OSDs are removed from the CRUSH map before deleting them. That is, be sure to do:

```
1. ceph osd crush rm osd.123
```

before:

```
1. ceph osd rm osd.123
```

## Snap Trimmer Improvements

This release greatly improves control and throttling of the snap trimmer. It introduces the “osd max trimming pgs” option (defaulting to 2), which limits how many PGs on an OSD can be trimming snapshots at a time. And it restores the safe use of the “osd snap trim sleep” option, which defaults to 0 but otherwise adds the given number of seconds in delay between every dispatch of trim operations to the underlying system.

## Other Notable Changes

- build/ops: “osd marked itself down” will not be recognised if host runs mon + osd on shutdown/reboot ([issue#18516](#), [pr#13492](#), Boris Ranto)
- build/ops: ceph-base package missing dependency for psmisc ([issue#19129](#), [pr#13786](#), Nathan Cutler)
- build/ops: enable build of ceph-resource-agents package on rpm-based os ([issue#17613](#), [issue#19546](#), [pr#13606](#), Nathan Cutler)
- build/ops: rbdmap.service not included in debian packaging (jewel-only) ([issue#19547](#), [pr#14383](#), Ken Dreyer)
- cephfs: Journaler may execute on\_safe contexts prematurely ([issue#20055](#), [pr#15468](#), “Yan, Zheng”)
- cephfs: MDS assert failed when shutting down ([issue#19204](#), [pr#14683](#), John Spray)

- cephfs: MDS goes readonly writing backtrace for a file whose data pool has been removed ([issue#19401](#), [pr#14682](#), John Spray)
- cephfs: MDS server crashes due to inconsistent metadata ([issue#19406](#), [pr#14676](#), John Spray)
- cephfs: No output for ceph mds rmfailed 0 -yes-i-really-mean-it command ([issue#16709](#), [pr#14674](#), John Spray)
- cephfs: Test failure: test\_data\_isolated  
(tasks.cephfs.test\_volume\_client.TestVolumeClient) ([issue#18914](#), [pr#14685](#), "Yan, Zheng")
- cephfs: Test failure: test\_open\_inode ([issue#18661](#), [pr#14669](#), John Spray)
- cephfs: The mount point break off when mds switch hanppened ([issue#19437](#), [pr#14679](#), Guan yunfei)
- cephfs: ceph-fuse does not recover after lost connection to MDS ([issue#16743](#), [issue#18757](#), [pr#14698](#), Kefu Chai, Henrik Korkuc, Patrick Donnelly)
- cephfs: client: fix the cross-quota rename boundary check conditions ([issue#18699](#), [pr#14667](#), Greg Farnum)
- cephfs: mds is crushed, after I set about 400 64KB xattr kv pairs to a file ([issue#19033](#), [pr#14684](#), Yang Honggang)
- cephfs: non-local quota changes not visible until some IO is done ([issue#17939](#), [pr#15466](#), John Spray, Nathan Cutler)
- cephfs: normalize file open flags internally used by cephfs ([issue#18872](#), [issue#19890](#), [pr#15000](#), Jan Fajerski, "Yan, Zheng")
- common: monitor creation with IPv6 public network segfaults ([issue#19371](#), [pr#14324](#), Fabian Grünbichler)
- common: radosstriper: protect aio\_write API from calls with 0 bytes ([issue#14609](#), [pr#13254](#), Sébastien Ponce)
- core: Objecter::epoch\_barrier isn't respected in \_op\_submit() ([issue#19396](#), [pr#14332](#), Ilya Dryomov)
- core: clear divergent\_priors set off disk ([issue#17916](#), [pr#14596](#), Greg Farnum)
- core: improve snap trimming, enable restriction of parallelism ([issue#19241](#), [pr#14492](#), Samuel Just, Greg Farnum)
- core: os/filestore/HashIndex: be loud about splits ([issue#18235](#), [pr#13788](#), Dan van der Ster)
- core: os/filestore: fix clang static check warn use-after-free ([issue#19311](#),

- pr#14044, liuchang0812, yaoning)
- core: transient jerasure unit test failures ([issue#18070](#), [issue#17762](#), [issue#18128](#), [issue#17951](#), [pr#14701](#), Kefu Chai, Pan Liu, Loic Dachary, Jason Dillaman)
  - core: two instances of omap\_digest mismatch ([issue#18533](#), [pr#14204](#), Samuel Just, David Zafman)
  - doc: Improvements to crushtool manpage ([issue#19649](#), [pr#14635](#), Loic Dachary, Nathan Cutler)
  - doc: PendingReleaseNotes: note about 19119 ([issue#19119](#), [pr#13732](#), Sage Weil)
  - doc: admin ops: fix the quota section ([issue#19397](#), [pr#14654](#), Chu, Hua-Rong)
  - doc: radosgw-admin: add the 'object stat' command to usage ([issue#19013](#), [pr#13872](#), Pavan Rallabhandi)
  - doc: rgw S3 create bucket should not do response in json ([issue#18889](#), [pr#13874](#), Abhishek Lekshmanan)
  - fs: Invalid error code returned by MDS is causing a kernel client WARNING ([issue#19205](#), [pr#13831](#), Jan Fajerski, xie xingguo)
  - librbd: Incomplete declaration for ContextWQ in librbd/Journal.h ([issue#18862](#), [pr#14152](#), Boris Ranto)
  - librbd: Issues with C API image metadata retrieval functions ([issue#19588](#), [pr#14666](#), Mykola Golub)
  - librbd: Possible deadlock performing a synchronous API action while refresh in-progress ([issue#18419](#), [pr#13154](#), Jason Dillaman)
  - librbd: is\_exclusive\_lock\_owner API should ping OSD ([issue#19287](#), [pr#14481](#), Jason Dillaman)
  - librbd: remove image header lock assertions ([issue#18244](#), [pr#13809](#), Jason Dillaman)
  - mds: C\_MDSInternalNoop::complete doesn't free itself ([issue#19501](#), [pr#14677](#), "Yan, Zheng")
  - mds: Too many stat ops when trying to probe a large file ([issue#19955](#), [pr#15472](#), "Yan, Zheng")
  - mds: avoid reusing deleted inode in StrayManager::\_purge\_stray\_logged ([issue#18877](#), [pr#14670](#), Zhi Zhang)
  - mds: enable start when session ino info is corrupt ([issue#19708](#), [issue#16842](#), [pr#14700](#), John Spray)

- mds: fragment space check can cause replayed request fail ([issue#18660](#), [pr#14668](#), "Yan, Zheng")
- mds: heartbeat timeout during rejoin, when working with large amount of caps/inodes ([issue#19118](#), [pr#14672](#), John Spray)
- mds: issue new caps when sending reply to client ([issue#19635](#), [pr#15438](#), "Yan, Zheng")
- mon: OSDMonitor: make 'osd crush move ...' work on osds ([issue#18587](#), [pr#13261](#), Sage Weil)
- mon: fix 'sortbitwise' warning on jewel ([issue#20578](#), [pr#15208](#), huanwen ren, Sage Weil)
- mon: make get\_mon\_log\_message() atomic ([issue#19427](#), [pr#14587](#), Kefu Chai)
- mon: remove bad rocksdb option ([issue#19392](#), [pr#14236](#), Sage Weil)
- msg: IPv6 Heartbeat packets are not marked with DSCP QoS - simple messenger ([issue#18887](#), [pr#13450](#), Yan Jun, Robin H. Johnson)
- msg: set close on exec flag ([issue#16390](#), [pr#13585](#), Kefu Chai)
- osd: -flush-journal: sporadic segfaults on exit ([issue#18820](#), [pr#13477](#), Alexey Sheplyakov)
- osd: Give requested scrubs a higher priority ([issue#15789](#), [pr#14686](#), David Zafman)
- osd: Implement asynchronous scrub sleep ([issue#19986](#), [issue#19497](#), [pr#15529](#), Brad Hubbard)
- osd: Object level shard errors are tracked and used if no auth available ([issue#20089](#), [pr#15416](#), David Zafman)
- osd: ReplicatedPG: try with pool's use-gmt setting if hitset archive not found ([issue#19185](#), [pr#13827](#), Kefu Chai)
- osd: allow client throttler to be adjusted on-fly, without restart ([issue#18791](#), [pr#13214](#), Piotr Dałek)
- osd: bypass readonly ops when osd full ([issue#19394](#), [pr#14181](#), Jianpeng Ma, yaoning)
- osd: degraded and misplaced status output inaccurate ([issue#18619](#), [pr#14325](#), David Zafman)
- osd: new added OSD always down when full flag is set ([issue#15025](#), [pr#14326](#), Mingxin Liu)
- osd: pg\_pool\_t::encode(): be compatible with Hammer <= 0.94.6 ([issue#19508](#),

pr#14392, Alexey Sheplyakov)

- osd: pre-jewel “osd rm” incrementals are misinterpreted ([issue#19119](#), [pr#13884](#), Ilya Dryomov)
- osd: preserve allocation hint attribute during recovery ([issue#19083](#), [pr#13647](#), yaoning)
- osd: promote throttle parameters are reversed ([issue#19773](#), [pr#14791](#), Mark Nelson)
- osd: reindex properly on pg log split ([issue#18975](#), [pr#14047](#), Alexey Sheplyakov)
- osd: restrict want\_acting to up+acting on recovery completion ([issue#18929](#), [pr#13541](#), Sage Weil)
- rbd-nbd: check /sys/block/nbdX/size to ensure kernel mapped correctly ([issue#18335](#), [pr#13932](#), Mykola Golub, Alexey Sheplyakov)
- rbd: [api] temporarily restrict (rbd\_)mirror\_peer\_add from adding multiple peers ([issue#19256](#), [pr#14664](#), Jason Dillaman)
- rbd: qemu crash triggered by network issues ([issue#18436](#), [pr#13244](#), Jason Dillaman)
- rbd: rbd -pool=x rename y z does not work ([issue#18326](#), [pr#14148](#), Gaurav Kumar Garg)
- rbd: systemctl stop rbdmap unmaps all rbds and not just the ones in /etc/ceph/rbdmap ([issue#18884](#), [issue#18262](#), [pr#14083](#), David Disseldorp, Nathan Cutler)
- rgw: “cluster [WRN] bad locator @X on object @X....” in cluster log ([issue#18980](#), [pr#14064](#), Casey Bodley)
- rgw: ‘radosgw-admin sync status’ on master zone of non-master zonegroup ([issue#18091](#), [pr#13779](#), Jing Wenjun)
- rgw: Change loglevel to 20 for ‘System already converted’ message ([issue#18919](#), [pr#13834](#), Vikhyat Umrao)
- rgw: Use decoded URI when verifying TempURL ([issue#18590](#), [pr#13724](#), Alexey Sheplyakov)
- rgw: a few cases where rgw\_obj is incorrectly initialized ([issue#19096](#), [pr#13842](#), Yehuda Sadeh)
- rgw: add apis to support ragweed suite ([issue#19804](#), [pr#14851](#), Yehuda Sadeh)
- rgw: add bucket size limit check to radosgw-admin ([issue#17925](#), [pr#14787](#), Matt Benjamin)

- rgw: allow system users to read SLO parts ([issue#19027](#), [pr#14752](#), Casey Bodley)
- rgw: don't return skew time in pre-signed url ([issue#18828](#), [issue#18829](#), [pr#14605](#), liuchang0812)
- rgw: failure to create s3 type subuser from admin rest api ([issue#16682](#), [pr#14815](#), snakeAngel2015)
- rgw: fix break inside of yield in RGWFetchAllMetaCR ([issue#17655](#), [pr#14066](#), Casey Bodley)
- rgw: fix failed to create bucket if a non-master zonegroup has a single zone ([issue#19756](#), [pr#14766](#), weiqiaomiao)
- rgw: health check errors out incorrectly ([issue#19025](#), [pr#13865](#), Pavan Rallabhandi)
- rgw: list\_plain\_entries() stops before bi\_log entries ([issue#19876](#), [pr#15383](#), Casey Bodley)
- rgw: multisite: fetch\_remote\_obj() gets wrong version when copying from remote ([issue#19599](#), [pr#14607](#), Zhang Shaowen, Casey Bodley)
- rgw: multisite: some yields in RGWMetaSyncShardCR::full\_sync() resume in incremental\_sync() ([issue#18076](#), [pr#13837](#), Casey Bodley, Abhishek Lekshmanan)
- rgw: only append zonegroups to rest params if not empty ([issue#20078](#), [pr#15312](#), Yehuda Sadeh, Karol Mroz)
- rgw: pullup civet chunked ([issue#19736](#), [pr#14776](#), Matt Benjamin)
- rgw: rgw\_file: fix event expire check, don't expire directories being read ([issue#19623](#), [issue#19270](#), [issue#19625](#), [issue#19624](#), [issue#19634](#), [issue#19435](#), [pr#14653](#), Gui Hecheng, Matt Benjamin)
- rgw: swift: disable revocation thread under certain circumstances ([issue#19499](#), [issue#9493](#), [pr#14789](#), Marcus Watts)
- rgw: the swift container acl does not support field .ref ([issue#18484](#), [pr#13833](#), Jing Wenjun)
- rgw: typo in rgw\_admin.cc ([issue#19026](#), [pr#13863](#), Ronak Jain)
- rgw: unsafe access in RGWListBucket\_ObjStore\_SWIFT::send\_response() ([issue#19249](#), [pr#14661](#), Yehuda Sadeh)
- rgw: upgrade to multisite v2 fails if there is a zone without zone info ([issue#19231](#), [pr#14136](#), Danny Al-Gaaf, Orit Wasserman)
- rgw: use separate http\_manager for read\_sync\_status ([issue#19236](#), [pr#14195](#), Casey Bodley, Shasha Lu)

- rgw: when converting region\_map we need to use rgw\_zone\_root\_pool ([issue#19195](#), [pr#14143](#), Orit Wasserman)
- rgw: zonegroupmap set does not work ([issue#19498](#), [issue#18725](#), [pr#14660](#), Orit Wasserman, Casey Bodley)
- rgw:fix memory leaks in data/md sync ([issue#20088](#), [pr#15382](#), weiqiaomiao)
- tests: 'ceph auth import -i' overwrites caps, should alert user before overwrite ([issue#18932](#), [pr#13544](#), Vikhyat Umrao)
- tests: New upgrade test for #19508 ([issue#19829](#), [issue#19508](#), [pr#14930](#), Nathan Cutler)
- tests: [ FAILED ] TestLibRBD.ImagePollIO in upgrade:client-upgrade-kraken-distro-basic-smithi ([issue#18617](#), [pr#13107](#), Jason Dillaman)
- tests: [ librados\_test\_stub] cls\_cxx\_map\_get\_XYZ methods don't return correct value ([issue#19597](#), [pr#14665](#), Jason Dillaman)
- tests: additional rbd-mirror test stability improvements ([issue#18935](#), [pr#14154](#), Jason Dillaman)
- tests: api\_misc: [ FAILED ] LibRadosMiscConnectFailure.ConnectFailure ([issue#15368](#), [pr#14763](#), Sage Weil)
- tests: buffer overflow in test LibCephFS.DirLs ([issue#18941](#), [pr#14671](#), "Yan, Zheng")
- tests: clone workunit using the branch specified by task ([issue#19429](#), [pr#14371](#), Kefu Chai, Dan Mick)
- tests: drop upgrade/hammer-jewel-x ([issue#20574](#), [pr#15933](#), Nathan Cutler)
- tests: dummy suite fails in OpenStack ([issue#18259](#), [pr#14070](#), Nathan Cutler)
- tests: eliminate race condition in Thrasher constructor ([issue#18799](#), [pr#13608](#), Nathan Cutler)
- tests: enable quotas for pre-luminous quota tests ([issue#20412](#), [pr#15936](#), Patrick Donnelly)
- tests: fix oversight in yaml comment ([issue#20581](#), [pr#14449](#), Nathan Cutler)
- tests: move swift.py task from teuthology to ceph, phase one (jewel) ([issue#20392](#), [pr#15870](#), Nathan Cutler, Sage Weil, Warren Usui, Greg Farnum, Ali Maredia, Tommi Virtanen, Zack Cerza, Sam Lang, Yehuda Sadeh, Joe Buck, Josh Durgin)
- tests: qa/Fixed upgrade sequence to 10.2.0 -> 10.2.7 -> latest -x (10.2.8) ([issue#20572](#), [pr#16089](#), Yuri Weinstein)

- tests: qa/suites/upgrade/hammer-x: set “sortbitwise” for jewel clusters ([issue#20342](#), [pr#15842](#), Nathan Cutler)
- tests: qa/workunits/rados/test-upgrade-\*: whitelist tests for master (part 1) ([issue#20577](#), [pr#15360](#), Sage Weil)
- tests: qa/workunits/rados/test-upgrade-\*: whitelist tests for master (part 2) ([issue#20576](#), [pr#15778](#), Kefu Chai)
- tests: qa/workunits/rados/test-upgrade-\*: whitelist tests the right way ([issue#20575](#), [pr#15824](#), Kefu Chai)
- tests: rados: sleep before ceph tell osd.0 flush\_pg\_stats after restart ([issue#16239](#), [issue#20489](#), [pr#14710](#), Kefu Chai, Nathan Cutler)
- tests: run upgrade/client-upgrade on latest CentOS 7.3 ([issue#20573](#), [pr#16088](#), Nathan Cutler)
- tests: run-rbd-unit-tests.sh assert in lockdep\_will\_lock, TestLibRBD.ObjectMapConsistentSnap ([issue#17447](#), [pr#14150](#), Jason Dillaman)
- tests: systemd test backport to jewel ([issue#19717](#), [pr#14694](#), Vasu Kulkarni)
- tests: test/librados/tmap\_migrate: g\_ceph\_context->put() upon return ([issue#20579](#), [pr#14809](#), Kefu Chai)
- tests: test\_notify.py: rbd.InvalidArgument: error updating features for image test\_notify\_clone2 ([issue#19692](#), [pr#14680](#), Jason Dillaman)
- tests: upgrade/hammer-x failing with OSD has the store locked when Thrasher runs ceph-objectstore-tool on down PG ([issue#19556](#), [pr#14416](#), Nathan Cutler)
- tests: upgrade:hammer-x/stress-split-erasure-code-x86\_64 fails in 10.2.8 integration testing ([issue#20413](#), [pr#15904](#), Nathan Cutler)
- tools: brag fails to count “in” mds ([issue#19192](#), [pr#14112](#), Oleh Prypin, Peng Zhang)
- tools: ceph-disk does not support cluster names different than ‘ceph’ ([issue#17821](#), [pr#14765](#), Loic Dachary)
- tools: ceph-disk: Racing between partition creation and device node creation ([issue#19428](#), [pr#14329](#), Erwan Velu)
- tools: ceph-disk: bluestore -setgroup incorrectly set with user ([issue#18955](#), [pr#13489](#), craigchi)
- tools: ceph-disk: ceph-disk list reports mount error for OSD having mount options with SELinux context ([issue#17331](#), [pr#14402](#), Brad Hubbard)
- tools: ceph-disk: do not setup\_statedir on trigger ([issue#19941](#), [pr#15504](#), Loic

Dachary)

- tools: ceph-disk: enable directory backed OSD at boot time ([issue#19628](#), [pr#14602](#), Loic Dachary)
- tools: rados: RadosImport::import should return an error if Rados::connect fails ([issue#19319](#), [pr#14113](#), Brad Hubbard)

## v10.2.7 Jewel

---

This point release fixes several important bugs in RBD mirroring, librbd & RGW.

We recommend that all v10.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- librbd: possible race in ExclusiveLock handle\_peer\_notification ([issue#19368](#), [pr#14233](#), Mykola Golub)
- osd: Increase priority for inactive PGs backfill ([issue#18350](#), [pr#13232](#), Bartłomiej Święcki)
- osd: Scrub improvements and other fixes ([issue#17857](#), [issue#18114](#), [issue#13937](#), [issue#18113](#), [pr#13146](#), Kefu Chai, David Zafman)
- osd: fix OSD network address in OSD heartbeat\_check log message ([issue#18657](#), [pr#13108](#), Vikhyat Umrao)
- rbd-mirror: deleting a snapshot during sync can result in read errors ([issue#18990](#), [pr#13596](#), Jason Dillaman)
- rgw: ‘period update’ does not remove short\_zone\_ids of deleted zones ([issue#15618](#), [pr#14140](#), Casey Bodley)
- rgw: DUMPABLE flag is cleared by setuid preventing core dumps ([issue#19089](#), [pr#13844](#), Brad Hubbard)
- rgw: clear data\_sync\_cr if RGWDataSyncControlCR fails ([issue#17569](#), [pr#13886](#), Casey Bodley)
- rgw: fix openssl ([issue#11239](#), [issue#19098](#), [issue#16535](#), [pr#14215](#), Marcus Watts)
- rgw: fix swift cannot disable object versioning with empty X-Versions-Location ([issue#18852](#), [pr#13823](#), Jing Wenjun)
- rgw: librgw: RGWLibFS::setattr fails on directories ([issue#18808](#), [pr#13778](#), Matt Benjamin)

- rgw: make sending Content-Length in 204 and 304 controllable ([issue#16602](#), [pr#13503](#), Radoslaw Zarzynski, Matt Benjamin)
- rgw: multipart uploads copy part support ([issue#12790](#), [pr#13219](#), Yehuda Sadeh, Javier M. Mellid, Matt Benjamin)
- rgw: multisite: RGWMetaSyncShardControlCR gives up on EIO ([issue#19019](#), [pr#13867](#), Casey Bodley)
- rgw: radosgw/swift: clean up flush / newline behavior ([issue#18473](#), [pr#14100](#), Nathan Cutler, Marcus Watts, Matt Benjamin)
- rgw: radosgw/swift: clean up flush / newline behavior. ([issue#18473](#), [pr#13143](#), Marcus Watts, Matt Benjamin)
- rgw: rgw\_fh: RGWFileHandle dtor must also cond-unlink from FHCache ([issue#19112](#), [pr#14231](#), Matt Benjamin)
- rgw: rgw\_file: avoid interning .. in FHCache table and don't ref for them ([issue#19036](#), [pr#13848](#), Matt Benjamin)
- rgw: rgw\_file: interned RGWFileHandle objects need parent refs ([issue#18650](#), [pr#13583](#), Matt Benjamin)
- rgw: rgw\_file: restore (corrected) fix for dir partial match (return of FLAG\_EXACT\_MATCH) ([issue#19060](#), [issue#18992](#), [issue#19059](#), [pr#13858](#), Matt Benjamin)
- rgw: rgw\_file: FHCache residence check should be exhaustive ([issue#19111](#), [pr#14169](#), Matt Benjamin)
- rgw: rgw\_file: ensure valid\_s3\_object\_name for directories, too ([issue#19066](#), [pr#13717](#), Matt Benjamin)
- rgw: rgw\_file: fix marker computation ([issue#19018](#), [issue#18989](#), [issue#18992](#), [issue#18991](#), [pr#13869](#), Matt Benjamin)
- rgw: rgw\_file: wip dir orphan ([issue#18992](#), [issue#18989](#), [issue#19018](#), [issue#18991](#), [pr#14205](#), Gui Hecheng, Matt Benjamin)
- rgw: rgw\_file: various fixes ([pr#14206](#), Matt Benjamin)
- rgw: rgw\_file: expand argv ([pr#14230](#), Matt Benjamin)

## v10.2.6 Jewel

---

This point release fixes several important bugs in RBD mirroring, RGW multi-site, CephFS, and RADOS.

We recommend that all v10.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

## OSDs No Longer Send ENXIO by Default

In previous versions, if a client sent an op to the wrong OSD, the OSD would reply with ENXIO. The rationale here is that the client or OSD is clearly buggy and we want to surface the error as clearly as possible. We now only send the ENXIO reply if the `osd_enxio_on_misdirected_op` option is enabled (it's off by default). This means that a VM using librbd that previously would have gotten an EIO and gone read-only will now see a blocked/hung IO instead.

## Other Notable Changes

- build/ops: add hostname sanity check to `run-{c}make-check.sh` ([issue#18134](#), [pr#12302](#), Nathan Cutler)
- build/ops: add ldap lib to rgw lib deps based on build config ([issue#17313](#), [pr#13183](#), Nathan Cutler)
- build/ops: ceph-create-keys loops forever ([issue#17753](#), [pr#11884](#), Alfredo Deza)
- build/ops: ceph daemons DUMPABLE flag is cleared by setuid preventing core dumps ([issue#17650](#), [pr#11736](#), Patrick Donnelly)
- build/ops: fixed compilation error when `-with-radowsgw=no` ([issue#18512](#), [pr#12729](#), Pan Liu)
- build/ops: fixed the issue when `-disable-server`, compilation fails. ([issue#18120](#), [pr#12239](#), Pan Liu)
- build/ops: fix undefined crypto references with `-with-xio` ([issue#18133](#), [pr#12296](#), Nathan Cutler)
- build/ops: install-deps.sh based on `/etc/os-release` ([issue#18466](#), [issue#18198](#), [pr#12405](#), Jan Fajerski, Nitin A Kamble, Nathan Cutler)
- build/ops: Remove the runtime dependency on `lsb_release` ([issue#17425](#), [pr#11875](#), John Coyle, Brad Hubbard)
- build/ops: rpm: `/etc/ceph/rbdmap` is packaged with executable access rights ([issue#17395](#), [pr#11855](#), Ken Dreyer)
- build/ops: selinux: Allow ceph to manage tmp files ([issue#17436](#), [pr#13048](#), Boris Ranto)
- build/ops: systemd: Restart Mon after 10s in case of failure ([issue#18635](#), [pr#13058](#), Wido den Hollander)
- build/ops: systemd restarts Ceph Mon to quickly after failing to start

- ([issue#18635](#), [pr#13184](#), Wido den Hollander)
- ceph-disk: fix flake8 errors ([issue#17898](#), [pr#11976](#), Ken Dreyer)
- cephfs: fuse client crash when adding a new osd ([issue#17270](#), [pr#11860](#), John Spray)
- cli: ceph-disk: convert none str to str before printing it ([issue#18371](#), [pr#13187](#), Kefu Chai)
- client: Fix lookup of “..” in jewel ([issue#18408](#), [pr#12766](#), Jeff Layton)
- client: fix stale entries in command table ([issue#17974](#), [pr#12137](#), John Spray)
- client: populate metadata during mount ([issue#18361](#), [pr#13085](#), John Spray)
- cli: implement functionality for adding, editing and removing omap values with binary keys ([issue#18123](#), [pr#12755](#), Jason Dillaman)
- common: Improve linux dcache hash algorithm ([issue#17599](#), [pr#11529](#), Yibo Cai)
- common: utime.h: fix timezone issue in round\_to\_\* funcs. ([issue#14862](#), [pr#11508](#), Zhao Chao)
- doc: Python Swift client commands in Quick Developer Guide don't match configuration in vstart.sh ([issue#17746](#), [pr#13043](#), Ronak Jain)
- librbd: allow to open an image without opening parent image ([issue#18325](#), [pr#13130](#), Ricardo Dias)
- librbd: metadata\_set API operation should not change global config setting ([issue#18465](#), [pr#13168](#), Mykola Golub)
- librbd: new API method to force break a peer's exclusive lock ([issue#15632](#), [issue#16773](#), [issue#17188](#), [issue#16988](#), [issue#17210](#), [issue#17251](#), [issue#18429](#), [issue#17227](#), [issue#18327](#), [issue#17015](#), [pr#12890](#), Danny Al-Gaaf, Mykola Golub, Jason Dillaman)
- librbd: properly order concurrent updates to the object map ([issue#16176](#), [pr#12909](#), Jason Dillaman)
- librbd: restore journal access when force disabling mirroring ([issue#17588](#), [pr#11916](#), Mykola Golub)
- mds: Cannot create deep directories when caps contain path=/somepath ([issue#17858](#), [pr#12154](#), Patrick Donnelly)
- mds: cephfs metadata pool: deep-scrub error omap\_digest != best guess omap\_digest ([issue#17177](#), [pr#12380](#), Yan, Zheng)
- mds: cephfs test failures (ceph.com/qa is broken, should be download.ceph.com/qa) ([issue#18574](#), [pr#13023](#), John Spray)

- mds: ceph-fuse crash during snapshot tests ([issue#18460](#), [pr#13120](#), Yan, Zheng)
- mds: ceph\_volume\_client: fix recovery from partial auth update ([issue#17216](#), [pr#11656](#), Ramana Raja)
- mds: ceph\_volume\_client.py : Error: Can't handle arrays of non-strings ([issue#17800](#), [pr#12325](#), Ramana Raja)
- mds: Cleanly reject session evict command when in replay ([issue#17801](#), [pr#12153](#), Yan, Zheng)
- mds: client segfault on ceph\_rmdir path / ([issue#9935](#), [pr#13029](#), Michal Jarzabek)
- mds: Clients without pool-changing caps shouldn't be allowed to change pool\_namespace ([issue#17798](#), [pr#12155](#), John Spray)
- mds: Decode errors on backtrace will crash MDS ([issue#18311](#), [pr#12836](#), Nathan Cutler, John Spray)
- mds: false failing to respond to cache pressure warning ([issue#17611](#), [pr#11861](#), Yan, Zheng)
- mds: finish clientreplay requests before requesting active state ([issue#18461](#), [pr#13113](#), Yan, Zheng)
- mds: fix incorrect assertion in Server::\_dir\_is\_nonempty() ([issue#18578](#), [pr#13459](#), Yan, Zheng)
- mds: fix MDSMap upgrade decoding ([issue#17837](#), [pr#13139](#), John Spray, Patrick Donnelly)
- mds: fix missing ll\_get for ll\_walk ([issue#18086](#), [pr#13125](#), Gui Hecheng)
- mds: Fix mount root for ceph\_mount users and change tarball format ([issue#18312](#), [issue#18254](#), [pr#12592](#), Jeff Layton)
- mds: fix null pointer dereference in Locker::handle\_client\_caps ([issue#18306](#), [pr#13060](#), Yan, Zheng)
- mds: lookup of ../ in returns -ENOENT ([issue#18408](#), [pr#12783](#), Jeff Layton)
- mds: MDS crashes on missing metadata object ([issue#18179](#), [pr#13119](#), Yan, Zheng)
- mds: mds fails to respawn if executable has changed ([issue#17531](#), [pr#11873](#), Patrick Donnelly)
- mds: MDS: false failing to respond to cache pressure warning ([issue#17716](#), [pr#11856](#), Yan, Zheng)
- mds: MDS goes damaged on blacklist (failed to read JournalPointer: -108 ((108) Cannot send after transport endpoint shutdown) ([issue#17236](#), [pr#11413](#), John Spray))

- mds: MDS long-time blocked ops. ceph-fuse locks up with getattr of file ([issue#17275](#), [pr#11858](#), Yan, Zheng)
- mds: speed up readdir by skipping unwanted dn ([issue#18519](#), [pr#12921](#), Xiaoxi Chen)
- mds: standby-replay daemons can sometimes miss events ([issue#17954](#), [pr#13126](#), John Spray)
- mon: cache tiering: base pool last\_force\_resend not respected (racing read got wrong version) ([issue#18366](#), [pr#13115](#), Sage Weil)
- mon: ceph osd down detection behaviour ([issue#18104](#), [pr#12677](#), xie xingguo)
- mon: Error EINVAL: removing mon.a at 172.21.15.16:6789/0, there will be 1 monitors ([issue#17725](#), [pr#11999](#), Joao Eduardo Luis)
- mon: health does not report pgs stuck in more than one state ([issue#17515](#), [pr#11660](#), Sage Weil)
- mon: monitor assertion failure when deactivating mds in (invalid) fscid 0 ([issue#17518](#), [pr#11862](#), Patrick Donnelly)
- mon: monitor cannot start because of FAILED assert(info.state == MDSMap::STATE\_STANDBY) ([issue#18166](#), [pr#13123](#), John Spray, Patrick Donnelly)
- mon: osd flag health message is misleading ([issue#18175](#), [pr#13117](#), Sage Weil)
- mon: OSDMonitor: clear jewel+ feature bits when talking to Hammer OSD ([issue#18582](#), [pr#13131](#), Piotr Dałek)
- mon: OSDs marked OUT wrongly after monitor failover ([issue#17719](#), [pr#11947](#), Dong Wu)
- mon: peon wrongly delete routed pg stats op before receive pg stats ack ([issue#18458](#), [pr#13045](#), Mingxin Liu)
- mon: send updated monmap to its subscribers ([issue#17558](#), [pr#11743](#), Kefu Chai)
- msgr: don't truncate message sequence to 32-bits ([issue#16122](#), [pr#12416](#), Yan, Zheng)
- msgr: msg/simple: clear\_pipe when wait() is mopping up pipes ([issue#15784](#), [pr#13062](#), Sage Weil)
- msgr: msg/simple/Pipe: error decoding addr ([issue#18072](#), [pr#12291](#), Sage Weil)
- osd: Add config option to disable new scrubs during recovery ([issue#17866](#), [pr#11944](#), Wido den Hollander)
- osd: collection\_list shadow return value # ([issue#17713](#), [pr#11737](#), Haomai Wang)

- osd: do not send ENXIO on misdirected op by default ([issue#18751](#), [pr#13255](#), Sage Weil)
- osd: FileStore: fiemap cannot be totally retrieved in xfs when the number of extents > 1364 ([issue#17610](#), [pr#11998](#), Kefu Chai, Ning Yao)
- osd: leveldb corruption leads to Operation not permitted not handled and assert ([issue#18037](#), [pr#12789](#), Nathan Cutler)
- osd: limit omap data in push op ([issue#16128](#), [pr#11991](#), Wanlong Gao)
- osd: osd crashes when radosgw-admin bi list -max-entries=1 command runing ([issue#17745](#), [pr#11758](#), weiqiaomiao)
- osd: osd\_max\_backfills default has changed, documentation should reflect that. ([issue#17701](#), [pr#11735](#), huangjun)
- osd: OSDMonitor: only reject MOSDBoot based on up\_from if inst matches ([issue#17899](#), [pr#12868](#), Samuel Just)
- osd: osd/PG: publish PG stats when backfill-related states change ([issue#18369](#), [pr#12875](#), Alexey Sheplyakov, Sage Weil)
- osd: Remove extra call to reg\_next\_scrub() during splits ([issue#16474](#), [pr#11606](#), David Zafman)
- osd: Revert "Merge pull request #12978 from asheplyakov/jewel-18581" ([issue#18809](#), [pr#13280](#), Samuel Just)
- osd: update\_log\_missing does not order correctly with osd\_ops ([issue#17789](#), [pr#11997](#), Samuel Just)
- qa/tasks: backport rbd\_fio fixes to jewel ([issue#13512](#), [pr#13104](#), Ilya Dryomov)
- qa/tasks/workunits: backport misc fixes to jewel ([issue#18336](#), [pr#12912](#), Sage Weil)
- rados: crash adding snap to purged\_snaps in ReplicatedPG::WaitingOnReplicas (part 2) ([issue#15943](#), [issue#18504](#), [pr#12791](#), Samuel Just)
- rados: Memory leaks in object\_list\_begin and object\_list\_end ([issue#18252](#), [pr#13118](#), Brad Hubbard)
- rados: The request lock RPC message might be incorrectly ignored ([issue#17030](#), [pr#10865](#), Jason Dillaman)
- rbd: add image id block name prefix APIs ([issue#18270](#), [pr#12529](#), Jason Dillaman)
- rbd: add max\_part and nbds\_max options in rbd nbd map, in order to keep consistent with ([issue#18186](#), [pr#12426](#), Pan Liu)
- rbd: Attempting to remove an image w/ incompatible features results in partial

- removal ([issue#18315](#), [pr#13156](#), Dongsheng Yang)
- rbd: bench-write will crash if -io-size is 4G ([issue#18422](#), [pr#13129](#), Gaurav Kumar Garg)
- rbd: diff calculate can hide parent extents when examining first snapshot in clone ([issue#18068](#), [pr#12322](#), Jason Dillaman)
- rbd: Exclusive lock improperly initialized on read-only image when using snap\_set API ([issue#17618](#), [pr#11852](#), Jason Dillaman)
- rbd: FAILED assert(m\_processing == 0) while running test\_lock\_fence.sh ([issue#17973](#), [pr#12323](#), Venky Shankar)
- rbd: Improve error reporting from rbd feature enable/disable ([issue#16985](#), [pr#13157](#), Gaurav Kumar Garg)
- rbd: JournalMetadata flooding with errors when being blacklisted ([issue#18243](#), [pr#12739](#), Jason Dillaman)
- rbd: librbd: use proper snapshot when computing diff parent overlap ([issue#18200](#), [pr#12649](#), Xiaoxi Chen)
- rbd: partition func should be enabled when load nbd.ko for rbd-nbd ([issue#18115](#), [pr#12754](#), Pan Liu)
- rbd: Potential race when removing two-way mirroring image ([issue#18447](#), [pr#13233](#), Mykola Golub)
- rbd: [qa] crash in journal-enabled fsx run ([issue#18618](#), [pr#13128](#), Jason Dillaman)
- rbd: 'rbd du' of missing image does not return error ([issue#16987](#), [pr#11854](#), Dongsheng Yang)
- rbd: rbd-mirror: gmock warnings in bootstrap request unit tests ([issue#18048](#), [issue#18012](#), [issue#18156](#), [issue#16991](#), [issue#18051](#), [pr#12425](#), Mykola Golub)
- rbd: rbd-mirror: image sync object map reload logs message ([issue#16179](#), [pr#12753](#), runsisi)
- rbd: rbd-mirror: snap protect of non-layered image results in split-brain ([issue#16962](#), [pr#11869](#), Mykola Golub)
- rbd: [rbd-mirror] sporadic image replayer shut down failure ([issue#18441](#), [pr#13155](#), Jason Dillaman)
- rbd: rbd-nbd: disallow mapping images >2TB in size ([issue#17219](#), [pr#11870](#), Mykola Golub)
- rbd: rbd-nbd: invalid error code for "failed to read nbd request" messages

([issue#18242](#), [pr#12756](#), Mykola Golub)

- rbd: status json format has duplicated/overwritten key ([issue#18261](#), [pr#12741](#), Mykola Golub)
- rbd: TestLibRBD.DiscardAfterWrite doesn't handle rbd\_skip\_partial\_discard = true ([issue#17750](#), [pr#11853](#), Jason Dillaman)
- rbd: truncate can cause unflushed snapshot data lose ([issue#17193](#), [pr#12324](#), Yan, Zheng)
- : ReplicatedBackend: take read locks for clone sources during recovery ([issue#17831](#), [issue#18583](#), [pr#12978](#), Samuel Just)
- rgw: add option to log custom HTTP headers (rgw\_log\_http\_headers) ([issue#18891](#), [pr#12490](#), Matt Benjamin)
- rgw: add support for Swift-at-root dependent features of Swift API ([issue#18526](#), [issue#16673](#), [pr#11497](#), Pritha Srivastava, Radoslaw Zarzynski, Pete Zaitcev, Abhishek Lekshmanan)
- rgw: add support for the prefix parameter in account listing of Swift API ([issue#17931](#), [pr#12258](#), Radoslaw Zarzynski)
- rgw: Add workaround for upgrade issues for older jewel versions ([issue#17820](#), [pr#12316](#), Orit Wasserman)
- rgw: be aware about tenants on cls\_user\_bucket -> rgw\_bucket conversion ([issue#18364](#), [issue#16355](#), [pr#13276](#), Radoslaw Zarzynski)
- rgw: bucket check remove \_multipart\_ prefix ([issue#13724](#), [pr#11470](#), Weijun Duan)
- rgw: bucket resharding ([issue#17549](#), [issue#17550](#), [pr#13341](#), Yehuda Sadeh, Robin H. Johnson)
- rgw: disable virtual hosting of buckets when no hostnames are configured ([issue#17440](#), [issue#15975](#), [issue#17136](#), [pr#11760](#), Casey Bodley, Robin H. Johnson)
- rgw: do not abort when accept a CORS request with short origin ([issue#18187](#), [pr#12397](#), LiuYang)
- rgw: don't store empty chains in gc ([issue#17897](#), [pr#12174](#), Yehuda Sadeh)
- rgw:fix for deleting objects name beginning and ending with underscores of one bucket using POST method of js sdk. ([issue#17888](#), [pr#12320](#), Casey Bodley)
- rgw: fix period update crash ([issue#18631](#), [pr#13273](#), Orit Wasserman)
- rgw: fix put\_acls for objects starting and ending with underscore ([issue#17625](#), [pr#11675](#), Orit Wasserman)
- rgw: fix use of marker in List::list\_objects() ([issue#18331](#), [pr#13358](#), Yehuda)

Sadeh)

- rgw: for the create\_bucket api, if the input creation\_time is zero, we ... ([issue#16597](#), [pr#11990](#), weiqiaomiao)
- rgw: Have a flavor of bucket deletion in radosgw-admin to bypass garbage collection ([issue#15557](#), [pr#10661](#), Pavan Rallabhandi)
- rgw: json encode/decode of RGWBucketInfo missing index\_type field ([issue#17755](#), [pr#11759](#), Yehuda Sadeh)
- rgw: ldap: enforce simple\_bind w/LDAPv3 redux ([issue#18339](#), [pr#12678](#), Weibing Zhang)
- rgw: leak from RGWMetaSyncShardCR::incremental\_sync ([issue#18412](#), [issue#18300](#), [pr#13004](#), Casey Bodley, Sage Weil)
- rgw: leak in RGWFetchAllMetaCR ([issue#17812](#), [pr#11872](#), Casey Bodley)
- rgw: librgw: objects created from s3 apis are not visible from nfs mount point ([issue#18651](#), [pr#13177](#), Matt Benjamin)
- rgw: log name instead of id for SystemMetaObj on failure ([issue#15776](#), [pr#12622](#), Wido den Hollander, Abhishek Lekshmanan)
- rgw: multimds: mds entering up:replay and processing down mds aborts ([issue#17670](#), [pr#11857](#), Patrick Donnelly)
- rgw: multipart upload copy ([issue#12790](#), [pr#13068](#), Yehuda Sadeh, Javier M. Mellid, Matt Benjamin)
- rgw: multisite: after finishing full sync on a bucket, incremental sync starts over from the beginning ([issue#17661](#), [issue#17624](#), [pr#11864](#), Zengran Zhang, Casey Bodley)
- rgw: multisite: assert(next) failed in RGWMetaSyncCR ([issue#17044](#), [pr#11477](#), Casey Bodley)
- rgw: multisite: coroutine deadlock assertion on error in FetchAllMetaCR ([issue#17571](#), [pr#11866](#), Casey Bodley)
- rgw: multisite: coroutine deadlock in RGWMetaSyncCR after ECANCELED errors ([issue#17465](#), [pr#12738](#), Casey Bodley)
- rgw: multisite doesn't retry RGWFetchAllMetaCR on failed lease ([issue#17047](#), [pr#11476](#), Casey Bodley)
- rgw: multisite: ECANCELED & 500 error on bucket delete ([issue#17698](#), [pr#12044](#), Casey Bodley)
- rgw: multisite: failed assertion in 'radosgw-admin bucket sync status'

- ([issue#18083](#), [pr#12314](#), Casey Bodley)
- rgw: multisite: fix ref counting of completions ([issue#17792](#), [issue#18414](#), [issue#17793](#), [issue#18407](#), [pr#13001](#), Casey Bodley)
  - rgw: multisite: metadata master can get the wrong value for 'oldest\_log\_period' ([issue#16894](#), [pr#11868](#), Casey Bodley)
  - rgw: multisite: obsolete 'radosgw-admin period prepare' command ([issue#17387](#), [pr#11574](#), Gaurav Kumar Garg)
  - rgw: multisite: race between ReadSyncStatus and InitSyncStatus leads to EIO errors ([issue#17568](#), [pr#11865](#), Casey Bodley)
  - rgw: multisite requests failing with '400 Bad Request' with civetweb 1.8 ([issue#17822](#), [pr#12313](#), Casey Bodley)
  - rgw: multisite: segfault after changing value of rgw\_data\_log\_num\_shards ([issue#18488](#), [pr#13180](#), Casey Bodley)
  - rgw: multisite: sync status reports master is on a different period ([issue#18064](#), [pr#13175](#), Abhishek Lekshmanan)
  - rgw: multisite upgrade from hammer -> jewel ignores rgw\_region\_root\_pool ([issue#17963](#), [pr#12156](#), Casey Bodley)
  - rgw: radosgw-admin period update reverts deleted zonegroup ([issue#17239](#), [pr#13171](#), Orit Wasserman)
  - rgw: Realm set does not create a new period ([issue#18333](#), [pr#13182](#), Orit Wasserman)
  - rgw: remove spurious mount entries for RGW buckets ([issue#17850](#), [pr#12045](#), Matt Benjamin)
  - rgw: Replacing '+' with "%20" in canonical uri for s3 v4 auth. ([issue#17076](#), [pr#12542](#), Pritha Srivastava)
  - rgw: rgw-admin: missing command to modify placement targets ([issue#18078](#), [pr#12428](#), Yehuda Sadeh, Casey Bodley)
  - rgw: RGWRados::get\_system\_obj() sends unnecessary stat request before read ([issue#17580](#), [pr#11867](#), Casey Bodley)
  - rgw: rgw\_rest\_s3: apply missed base64 try-catch ([issue#17663](#), [pr#11672](#), Matt Benjamin)
  - rgw: RGW will not list Argonaut-era bucket via HTTP (but radosgw-admin works) ([issue#17372](#), [pr#11863](#), Yehuda Sadeh)
  - rgw: sends omap\_getvals with (u64)-1 limit ([issue#17985](#), [pr#12419](#), Yehuda Sadeh,

Sage Weil)

- rgw: slave zonegroup cannot enable the bucket versioning ([issue#18003](#), [pr#13173](#), Orit Wasserman)
- rgw: TempURL properly handles accounts created with the implicit tenant ([issue#17961](#), [pr#12079](#), Radoslaw Zarzynski)
- rgw: the value of total\_time is wrong in the result of 'radosgw-admin log show' opt ([issue#17598](#), [pr#11876](#), weiqiaomiao)
- rgw: Unable to commit period zonegroup change ([issue#17364](#), [pr#12315](#), Orit Wasserman)
- rgw: valgrind "invalid read size 4" RGWGetObj ([issue#18071](#), [pr#12997](#), Matt Benjamin)
- rgw: work around curl\_multi\_wait bug with non-blocking reads ([issue#15915](#), [issue#16368](#), [issue#16695](#), [pr#11627](#), John Coyle, Casey Bodley)
- tests: add require\_jewel\_osds before upgrading last hammer node ([issue#18719](#), [pr#13161](#), Nathan Cutler)
- tests: add require\_jewel\_osds to upgrade/hammer-x/tiering ([issue#18920](#), [pr#13404](#), Nathan Cutler)
- tests: assertion failure in a radosgw-admin related task ([issue#17167](#), [pr#12764](#), Orit Wasserman)
- tests: Cannot reserve CentOS 7.2 smithi machines ([issue#18416](#), [issue#18401](#), [pr#13050](#), Nathan Cutler, Sage Weil, Yuri Weinstein)
- tests: ignore bogus ceph-objectstore-tool error in ceph\_manager ([issue#16263](#), [pr#13240](#), Nathan Cutler, Kefu Chai)
- tests: objecter\_requests workunit fails on wip branches ([issue#18393](#), [pr#12761](#), Sage Weil)
- tests: qa/suites/upgrade/hammer-x: break stress split ec symlinks ([issue#19006](#), [pr#13533](#), Nathan Cutler)
- tests: qa/suites/upgrade/hammer-x/stress-split: finish thrashing before final upgrade ([issue#19004](#), [pr#13222](#), Sage Weil)
- tests: qa/tasks/ceph\_deploy.py: use dev option ([issue#18736](#), [pr#13106](#), Vasu Kulkarni)
- tests: qa/workunits/rbd: use more recent qemu-io tests that support Xenial ([issue#18149](#), [issue#10773](#), [pr#13103](#), Jason Dillaman)
- tests: remove qa/suites/buildpackages ([issue#18846](#), [pr#13299](#), Loic Dachary)

- tests: SUSE yaml facets in qa/distros/all are out of date ([issue#18856](#), [issue#18846](#), [pr#13331](#), Nathan Cutler)
- tests: update rbd/singleton/all/formatted-output.yaml to support ceph-ci ([issue#18440](#), [pr#12822](#), Nathan Cutler, Venky Shankar)
- tests: update Ubuntu image url after ceph.com refactor ([issue#18542](#), [pr#12959](#), Jason Dillaman)
- tests: upgrade:hammer-x: install firefly only on Ubuntu 14.04 ([issue#18089](#), [pr#13153](#), Nathan Cutler)
- tests: use ceph-jewel branch for s3tests ([issue#18384](#), [pr#12745](#), Nathan Cutler)
- tests: Workunits needlessly wget from git.ceph.com ([issue#18336](#), [issue#18271](#), [issue#18388](#), [pr#12686](#), Nathan Cutler, Sage Weil)
- test: temporarily disable fork()'ing tests ([issue#16556](#), [issue#17832](#), [pr#11953](#), John Spray)
- test: test fails due to The UNIX domain socket path ([issue#16014](#), [pr#12151](#), Loic Dachary)
- tools: ceph-disk: ceph-disk@.service races with ceph-osd@.service ([issue#17889](#), [issue#17813](#), [pr#12147](#), Loic Dachary)
- tools: ceph-disk -dmcrypt create must not require admin key ([issue#17849](#), [pr#12033](#), Loic Dachary)
- tools: ceph-disk prepare writes osd log 0 with root owner ([issue#18538](#), [pr#13025](#), Samuel Matzek)
- tools: crushtool -compile is create output despite of missing item ([issue#17306](#), [pr#11410](#), Kefu Chai)
- tools: rados bench seq must verify the hostname ([issue#17526](#), [pr#13049](#), Loic Dachary)
- tools: snapshotted RBD extent objects can't be manually evicted from a cache tier ([issue#17896](#), [pr#11968](#), Mingxin Liu)
- tools: systemd/ceph-disk: reduce ceph-disk flock contention ([issue#18049](#), [issue#13160](#), [pr#12210](#), David Disseldorp)

## v10.2.5 Jewel

---

This point release fixes an important [regression introduced in v10.2.4](#).

We recommend that all v10.2.x users upgrade.

## Notable Changes

---

For more detailed information, see [the complete changelog](#).

- msg/simple/Pipe: avoid returning 0 on poll timeout ([issue#18185](#), [pr#12376](#), Sage Weil)

## v10.2.4 Jewel

---

This point release fixes several important bugs in RBD mirroring, RGW multi-site, CephFS, and RADOS.

We recommend that all v10.2.x users upgrade. Also note the following when upgrading from hammer

## Upgrading from hammer

---

When the last hammer OSD in a cluster containing jewel MONs is upgraded to jewel, as of 10.2.4 the jewel MONs will issue this warning: “all OSDs are running jewel or later but the ‘require\_jewel\_osds’ osdmap flag is not set” and change the cluster health status to HEALTH\_WARN.

This is a signal for the admin to do “ceph osd set require\_jewel\_osds” - by doing this, the upgrade path is complete and no more pre-Jewel OSDs may be added to the cluster.

## Notable Changes

---

For more detailed information, see [the complete changelog](#).

- build/ops: aarch64: Compiler-based detection of crc32 extended CPU type is broken ([issue#17516](#), [pr#11492](#), Alexander Graf)
- build/ops: allow building RGW with LDAP disabled ([issue#17312](#), [pr#11478](#), Daniel Gryniewicz)
- build/ops: backport ‘logrotate: Run as root/ceph’ ([issue#17381](#), [pr#11201](#), Boris Ranto)
- build/ops: ceph installs stuff in %\_udevrulesdir but does not own that directory ([issue#16949](#), [pr#10862](#), Nathan Cutler)
- build/ops: ceph-osd-prestart.sh fails confusingly when data directory does not exist ([issue#17091](#), [pr#10812](#), Nathan Cutler)

- build/ops: disable LTTng-UST in openSUSE builds ([issue#16937](#), [pr#10794](#), Michel Normand)
- build/ops: i386 tarball gitbuilder failure on master ([issue#16398](#), [pr#10855](#), Vikhyat Umrao, Kefu Chai)
- build/ops: include more files in “make dist” tarball ([issue#17560](#), [pr#11431](#), Ken Dreyer)
- build/ops: incorrect value of CINIT\_FLAG\_DEFER\_DROP\_PRIVILEGES ([issue#16663](#), [pr#10278](#), Casey Bodley)
- build/ops: remove SYSTEMD\_RUN from initscript ([issue#7627](#), [issue#16441](#), [issue#16440](#), [pr#9872](#), Vladislav Odintsov)
- build/ops: systemd: add install section to rbdmap.service file ([issue#17541](#), [pr#11158](#), Jelle vd Kooij)
- common: Enable/Disable of features is allowed even the features are already enabled/disabled ([issue#16079](#), [pr#11460](#), Lu Shi)
- common: Log.cc: Assign LOG\_INFO priority to syslog calls ([issue#15808](#), [pr#11231](#), Brad Hubbard)
- common: Proxied operations shouldn't result in error messages if replayed ([issue#16130](#), [pr#11461](#), Vikhyat Umrao)
- common: Request exclusive lock if owner sends -ENOTSUPP for proxied maintenance op ([issue#16171](#), [pr#10784](#), Jason Dillaman)
- common: msgr/async: Messenger thread long time lock hold risk ([issue#15758](#), [pr#10761](#), Wei Jin)
- doc: fix description for rsize and rasize ([issue#17357](#), [pr#11171](#), Andreas Gerstmayr)
- filestore: can get stuck in an unbounded loop during scrub ([issue#17859](#), [pr#12001](#), Sage Weil)
- fs: Failure in snaptest-git-ceph.sh ([issue#17172](#), [pr#11419](#), Yan, Zheng)
- fs: Log path as well as ino when detecting metadata damage ([issue#16973](#), [pr#11418](#), John Spray)
- fs: client: FAILED assert(root\_ancestor->qtree == \_\_null) ([issue#16066](#), [issue#16067](#), [pr#10107](#), Yan, Zheng)
- fs: client: add missing client\_lock for get\_root ([issue#17197](#), [pr#10921](#), Patrick Donnelly)
- fs: client: fix shutdown with open inodes ([issue#16764](#), [pr#10958](#), John Spray)

- fs: client: nlink count is not maintained correctly ([issue#16668](#), [pr#10877](#), Jeff Layton)
- fs: multimds: allow\_multimds not required when max\_mds is set in ceph.conf at startup ([issue#17105](#), [pr#10997](#), Patrick Donnelly)
- librados: memory leaks from ceph::crypto (WITH\_NSS) ([issue#17205](#), [pr#11409](#), Casey Bodley)
- librados: modify Pipe::connect() to return the error code ([issue#15308](#), [pr#11193](#), Vikhyat Umrao)
- librados: remove new setxattr overload to avoid breaking the C++ ABI ([issue#18058](#), [pr#12207](#), Josh Durgin)
- librbd: cannot disable journaling or remove non-mirrored, non-primary image ([issue#16740](#), [pr#11337](#), Jason Dillaman)
- librbd: discard after write can result in assertion failure ([issue#17695](#), [pr#11644](#), Jason Dillaman)
- librbd::Operations: update notification failed: (2) No such file or directory ([issue#17549](#), [pr#11420](#), Jason Dillaman)
- mds: Crash in Client::\_invalidate\_kernel\_dcache when reconnecting during unmount ([issue#17253](#), [pr#11414](#), Yan, Zheng)
- mds: Duplicate damage table entries ([issue#17173](#), [pr#11412](#), John Spray)
- mds: Failure in dirfrag.sh ([issue#17286](#), [pr#11416](#), Yan, Zheng)
- mds: Failure in snaptest-git-ceph.sh ([issue#17271](#), [pr#11415](#), Yan, Zheng)
- mon: Ceph Status - Segmentation Fault ([issue#16266](#), [pr#11408](#), Brad Hubbard)
- mon: Display full flag in ceph status if full flag is set ([issue#15809](#), [pr#9388](#), Vikhyat Umrao)
- mon: Error EINVAL: removing mon.a at 172.21.15.16:6789/0, there will be 1 monitors ([issue#17725](#), [pr#12267](#), Joao Eduardo Luis)
- mon: OSDMonitor: only reject MOSDBoot based on up\_from if inst matches ([issue#17899](#), [pr#12067](#), Samuel Just)
- mon: OSDMonitor: Missing nearfull flag set ([issue#17390](#), [pr#11272](#), Igor Podoski)
- mon: Upgrading 0.94.6 -> 0.94.9 saturating mon node networking ([issue#17365](#), [issue#17386](#), [pr#11679](#), Sage Weil, xie xingguo)
- mon: ceph mon Segmentation fault after set crush\_ruleset ceph 10.2.2 ([issue#16653](#), [pr#10861](#), song baisen)

- mon: crash: crush/CrushWrapper.h: 940: FAILED assert(successful\_detach) ([issue#16525](#), [pr#10496](#), Kefu Chai)
- mon: don't crash on invalid standby\_for\_fscid ([issue#17466](#), [pr#11389](#), John Spray)
- mon: fix missing osd metadata (again) ([issue#17685](#), [pr#11642](#), John Spray)
- mon: osdmonitor: decouple adjust\_heartbeat\_grace and min\_down\_reporters ([issue#17055](#), [pr#10757](#), Zengran Zhang)
- mon: the %USED of ceph df is wrong ([issue#16933](#), [pr#10860](#), Kefu Chai)
- osd: condition OSDMap encoding on features ([issue#18015](#), [pr#12167](#), Sage Weil)
- osd: PG:::\_update\_calc\_stats wrong for CRUSH\_ITEM\_NONE up set items ([issue#16998](#), [pr#10883](#), Samuel Just)
- osd: PG::choose\_acting valgrind error or ./common/hobject.h: 182: FAILED assert(!max || (\*this == hobject\_t(hobject\_t::get\_max()))) ([issue#13967](#), [pr#10885](#), Tao Chang)
- osd: Potential crash during journal::Replay shut down ([issue#16433](#), [pr#10645](#), Jason Dillaman)
- osd: add peer\_addr in heartbeat\_check log message ([issue#15762](#), [pr#9739](#), Vikhyat Umrao, Sage Weil)
- osd: adjust scrub boundary to object without SnapSet ([issue#17470](#), [pr#11311](#), Samuel Just)
- osd: ceph osd df does not show summarized info correctly if one or more OSDs are out ([issue#16706](#), [pr#10759](#), xie xingguo)
- osd: journal: do not prematurely flag object recorder as closed ([issue#17590](#), [pr#11634](#), Jason Dillaman)
- osd: mark\_all\_unfound\_lost() leaves unapplied changes ([issue#16156](#), [pr#10886](#), Samuel Just)
- osd: segfault in ObjectCacher::FlusherThread ([issue#16610](#), [pr#10864](#), Yan, Zheng)
- qa: remove EnumerateObjects from librados upgrade tests ([pr#11728](#), Josh Durgin)
- rbd: Disabling pool mirror mode with registered peers results orphaned mirrored images ([issue#16984](#), [pr#10857](#), Jason Dillaman)
- rbd: Imagewatcher: use after free within C\_UnwatchAndFlush ([issue#17289](#), [issue#17254](#), [pr#11466](#), Jason Dillaman)
- rbd: Prevent the creation of a clone from a non-primary mirrored image ([issue#16449](#), [pr#10650](#), Mykola Golub)

- rbd: RBD should restrict mirror enable/disable actions on parents/clones ([issue#16056](#), [pr#11459](#), zhuangzeqiang)
- rbd: TestJournalReplay: sporadic assert(`m_state == STATE_READY || m_state == STATE_STOPPING`) failure ([issue#17566](#), [pr#11590](#), Jason Dillaman)
- rbd: bench io-size should not be larger than image size ([issue#16967](#), [pr#10796](#), Jason Dillaman)
- rbd: ceph 10.2.2 rbd status on image format 2 returns (2) No such file or directory ([issue#16887](#), [pr#10652](#), Jason Dillaman)
- rbd: helgrind: TestLibRBD.TestIOPP potential deadlock closing an image with read-ahead enabled ([issue#17198](#), [pr#11463](#), Jason Dillaman)
- rbd: image.stat() call in librbdpy fails sometimes ([issue#17310](#), [pr#11464](#), Jason Dillaman)
- rbd: krbd qa scripts and concurrent.sh test fix ([issue#17223](#), [pr#11018](#), Ilya Dryomov)
- rbd: krbd-related CLI patches ([issue#17554](#), [pr#11400](#), Ilya Dryomov)
- rbd: mirror: improve resiliency of stress test case ([issue#16855](#), [issue#16555](#), [issue#14738](#), [issue#15259](#), [issue#17446](#), [issue#17355](#), [issue#16538](#), [issue#16974](#), [issue#17283](#), [issue#17317](#), [issue#17416](#), [issue#16227](#), [pr#11433](#), Mykola Golub, Ricardo Dias, Jason Dillaman)
- rbd: rbd-nbd IO hang ([issue#16921](#), [pr#11467](#), Jason Dillaman)
- rbd: update\_features API needs to support backwards/forward compatibility ([issue#17330](#), [pr#11462](#), Jason Dillaman)
- rgw: COPY broke multipart files uploaded under dumpling ([issue#16435](#), [pr#10866](#), Yehuda Sadeh)
- rgw: Config parameter rgw keystone make new tenants in radosgw multitenancy does not work ([issue#17293](#), [pr#11473](#), SirishaGuduru)
- rgw: Do not archive metadata by default ([issue#17256](#), [pr#11321](#), Pavan Rallabhandi, Matt Benjamin)
- rgw: ERROR: got unexpected error when trying to read object: -2 ([issue#17111](#), [pr#11472](#), Yang Honggang)
- rgw: Modification for TEST S3 ACCESS section in INSTALL CEPH OBJECT GATEWAY page ([issue#15603](#), [pr#11475](#), la-sguduru)
- rgw: RGW loses realm/period/zonegroup/zone data: period overwritten if somewhere in the cluster is still running Hammer ([issue#17371](#), [pr#11519](#), Orit Wasserman)

- rgw: RGWDataSyncCR fails on errors from RGWListBucketIndexesCR ([issue#17073](#), [pr#11330](#), Casey Bodley)
- rgw: S3 object versioning fails when applied on a non-master zone ([issue#16494](#), [pr#11367](#), Yehuda Sadeh)
- rgw: add orphan options to radosgw-admin -help and man page ([issue#17281](#), [issue#17280](#), [pr#11139](#), Ken Dreyer, Thomas Serlin)
- rgw: back off bucket sync on failures, don't store marker ([issue#16742](#), [pr#11021](#), Yehuda Sadeh)
- rgw: combined LDAP backports ([issue#17544](#), [issue#17185](#), [pr#11332](#), Harald Klein, Matt Benjamin)
- rgw: cors auto memleak ([issue#16564](#), [pr#10656](#), Yan Jun)
- rgw: default quota fixes ([issue#16410](#), [pr#10832](#), Pavan Rallabhandi, Daniel Gryniewicz)
- rgw: doc: description of multipart part entity is wrong ([issue#17504](#), [pr#11342](#), weiqiaomiao)
- rgw: don't loop forever when reading data from 0 sized segment. ([issue#17692](#), [pr#11626](#), Marcus Watts)
- rgw: fix put\_acls for objects starting and ending with underscore ([issue#17625](#), [pr#11669](#), Orit Wasserman)
- rgw: fix regression with handling double underscore ([issue#17443](#), [issue#16856](#), [pr#11563](#), Yehuda Sadeh, Orit Wasserman)
- rgw: handle empty POST condition ([issue#17635](#), [pr#11662](#), Yehuda Sadeh)
- rgw: metadata sync can skip markers for failed/incomplete entries ([issue#16759](#), [pr#10657](#), Yehuda Sadeh)
- rgw: nfs backports ([issue#17393](#), [issue#17311](#), [issue#17367](#), [issue#17319](#), [issue#17321](#), [issue#17322](#), [issue#17323](#), [issue#17325](#), [issue#17326](#), [issue#17327](#), [pr#11335](#), Min Chen, Yan Jun, Weibing Zhang, Matt Benjamin)
- rgw: period commit loses zonegroup changes: region\_map converted repeatedly ([issue#17051](#), [pr#10890](#), Casey Bodley)
- rgw: period commit return error when the current period has a zonegroup which doesn't have a master zone ([issue#17110](#), [pr#10867](#), weiqiaomiao)
- rgw: radosgw daemon core when reopen logs ([issue#17036](#), [pr#10868](#), weiqiaomiao)
- rgw: rgw file uses too much CPU in gc/idle thread ([issue#16976](#), [pr#10889](#), Matt Benjamin)

- rgw: s3tests-test-readwrite failing with 500 ([issue#16930](#), [pr#11471](#), Yehuda Sadeh)
- rgw: upgrade from old multisite to new multisite fails ([issue#16751](#), [pr#10891](#), Orit Wasserman)
- rgw: response information is error when getting token of swift account ([issue#15195](#), [pr#11474](#), Qiankun Zheng)
- rgw: user email can modify to empty when it has values ([issue#13286](#), [pr#11469](#), Yehuda Sadeh, Weijun Duan)
- tests: ceph-disk must ignore debug monc ([issue#17607](#), [pr#11548](#), Loic Dachary)
- tests: fix TestClsRbd.mirror\_image failure in upgrade:jewel-x-master-distro-basic-vps ([issue#16529](#), [pr#10888](#), Jason Dillaman)
- tests: scsi\_debug fails /dev/disk/by-partuuid ([issue#17100](#), [pr#11411](#), Loic Dachary)
- tests: test/ceph\_test\_msgr: do not use Message::middle for holding transient... ([issue#17365](#), [issue#17728](#), [issue#16955](#), [pr#11742](#), Haomai Wang, Kefu Chai, Michal Jarzabek, Sage Weil)
- tools: Missing comma in ceph-create-keys causes concatenation of arguments ([issue#17815](#), [pr#11822](#), Patrick Donnelly)
- tools: add a tool to rebuild mon store from OSD ([issue#17179](#), [issue#17400](#), [pr#11126](#), Kefu Chai, xie xingguo)
- tools: ceph-create-keys: sometimes blocks forever if mds allow is set ([issue#16255](#), [pr#11417](#), John Spray)
- tools: ceph-disk should timeout when a lock cannot be acquired ([issue#16580](#), [pr#10758](#), Loic Dachary)
- tools: ceph-disk: expected systemd unit failures are confusing ([issue#15990](#), [pr#10884](#), Boris Ranto)
- tools: ceph-disk: using a regular file as a journal fails ([issue#16280](#), [issue#17662](#), [pr#11657](#), Jayashree Candadai, Anirudha Bose, Loic Dachary, Shylesh Kumar)
- tools: ceph-objectstore-tool crashes if -journal-path <a-directory> ([issue#17307](#), [pr#11407](#), Kefu Chai)
- tools: ceph-objectstore-tool: add a way to split filestore directories offline ([issue#17220](#), [pr#11252](#), Josh Durgin)
- tools: ceph-post-file: use new ssh key ([issue#14267](#), [pr#11746](#), David Galloway)

## v10.2.3 Jewel

This point release fixes several important bugs in RBD mirroring, RGW multi-site, CephFS, and RADOS.

We recommend that all v10.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

- build/ops: 60-ceph-partuuid-workaround-rules still needed by debian jessie (udev 215-17) ([issue#16351](#), [pr#10653](#), runsisi, Loic Dachary)
- build/ops: ceph Resource Agent does not work with systemd ([issue#14828](#), [pr#9917](#), Nathan Cutler)
- build/ops: ceph-base requires parted ([issue#16095](#), [pr#10008](#), Ken Dreyer)
- build/ops: ceph-osd-prestart.sh contains Upstart-specific code ([issue#15984](#), [pr#10364](#), Nathan Cutler)
- build/ops: mount.ceph: move from ceph-base to ceph-common and add symlink in /sbin for SUSE ([issue#16598](#), [issue#16645](#), [pr#10357](#), Nathan Cutler, Dan Horák, Ricardo Dias, Kefu Chai)
- build/ops: need rocksdb commit 7ca731b12ce for ppc64le build ([issue#17092](#), [pr#10816](#), Nathan Cutler)
- build/ops: rpm: OBS needs ExclusiveArch ([issue#16936](#), [pr#10614](#), Michel Normand)
- cli: ceph command line tool chokes on ceph -w (the dash is unicode ‘en dash’ &ndash, copy-paste to reproduce) ([issue#12287](#), [pr#10420](#), Oleh Prypin, Kefu Chai)
- common: expose buffer const\_iterator symbols ([issue#16899](#), [pr#10552](#), Noah Watkins)
- common: global-init: fixup chown of the run directory along with log and asok files ([issue#15607](#), [pr#8754](#), Karol Mroz)
- fs: ceph-fuse: link to libtcmalloc or jemalloc ([issue#16655](#), [pr#10303](#), Yan, Zheng)
- fs: client: crash in unmount when fuse\_use\_invalidate\_cb is enabled ([issue#16137](#), [pr#10106](#), Yan, Zheng)
- fs: client: fstat cap release ([issue#15723](#), [pr#9562](#), Yan, Zheng, Noah Watkins)
- fs: essential backports for OpenStack Manila ([issue#15406](#), [issue#15614](#), [issue#15615](#), [pr#10453](#), John Spray, Ramana Raja, Xiaoxi Chen)

- fs: fix double-unlock on shutdown ([issue#17126](#), [pr#10847](#), Greg Farnum)
- fs: fix mdsmap print\_summary with standby replays ([issue#15705](#), [pr#9547](#), John Spray)
- fs: fuse mounted file systems fails SAMBA CTDB ping\_pong rw test with v9.0.2 ([issue#12653](#), [issue#15634](#), [pr#10108](#), Yan, Zheng)
- librados: Missing export for rados\_aio\_get\_version in src/include/rados/librados.h ([issue#15535](#), [pr#9574](#), Jim Wright)
- librados: osd: bad flags can crash the osd ([issue#16012](#), [pr#9997](#), Sage Weil)
- librbd: Close journal and object map before flagging exclusive lock as released ([issue#16450](#), [pr#10053](#), Jason Dillaman)
- librbd: Crash when utilizing advisory locking API functions ([issue#16364](#), [pr#10051](#), Jason Dillaman)
- librbd: ExclusiveLock object leaked when switching to snapshot ([issue#16446](#), [pr#10054](#), Jason Dillaman)
- librbd: FAILED assert(object\_no < m\_object\_map.size()) ([issue#16561](#), [pr#10647](#), Jason Dillaman)
- librbd: Image removal doesn't necessarily clean up all rbd\_mirroring entries ([issue#16471](#), [pr#10009](#), Jason Dillaman)
- librbd: Object map/fast-diff invalidated if journal replays the same snap remove event ([issue#16350](#), [pr#10010](#), Jason Dillaman)
- librbd: Timeout sending mirroring notification shouldn't result in failure ([issue#16470](#), [pr#10052](#), Jason Dillaman)
- librbd: Whitelist EBUSY error from snap unprotect for journal replay ([issue#16445](#), [pr#10055](#), Jason Dillaman)
- librbd: cancel all tasks should wait until finisher is done ([issue#16517](#), [pr#9752](#), Haomai Wang)
- librbd: delay acquiring lock if image watch has failed ([issue#16923](#), [pr#10827](#), Jason Dillaman)
- librbd: fix missing return statement if failed to get mirror image state ([issue#16600](#), [pr#10144](#), runsisi)
- librbd: flag image as updated after proxying maintenance op ([issue#16404](#), [pr#9883](#), Jason Dillaman)
- librbd: mkfs.xfs slow performance with discards and object map ([issue#16707](#), [issue#16689](#), [pr#10649](#), Jason Dillaman)

- librbd: potential use after free on refresh error ([issue#16519](#), [pr#9952](#), Mykola Golub)
- librbd: rbd-nbd does not properly handle resize notifications ([issue#15715](#), [pr#10679](#), Mykola Golub)
- librbd: the option ‘rbd\_cache\_writethrough\_until\_flush=true’ doesn’t work ([issue#16740](#), [issue#16386](#), [issue#16708](#), [issue#16654](#), [issue#16478](#), [pr#10797](#), Mykola Golub, xinxin shu, Xiaowei Chen, Jason Dillaman)
- mds: tell command blocks forever with async messenger (TestVolumeClient.test\_evict\_client failure) ([issue#16288](#), [pr#10501](#), Douglas Fuller)
- mds: Confusing MDS log message when shut down with stalled journaler reads ([issue#15689](#), [pr#9557](#), John Spray)
- mds: Deadlock on shutdown active rank while busy with metadata IO ([issue#16042](#), [pr#10502](#), Patrick Donnelly)
- mds: Failing file operations on kernel based cephfs mount point leaves unaccessible file behind on hammer 0.94.7 ([issue#16013](#), [pr#10199](#), Yan, Zheng)
- mds: Fix shutting down mds timed-out due to deadlock ([issue#16396](#), [pr#10500](#), Zhi Zhang)
- mds: MDSMonitor fixes ([issue#16136](#), [pr#9561](#), xie xingguo)
- mds: MDSMonitor::check\_subs() is very buggy ([issue#16022](#), [pr#10103](#), Yan, Zheng)
- mds: Session::check\_access() is buggy ([issue#16358](#), [pr#10105](#), Yan, Zheng)
- mds: StrayManager.cc: 520: FAILED assert(dnl->is\_primary()) ([issue#15920](#), [pr#9559](#), Yan, Zheng)
- mds: enforce a dirfrag limit on entries ([issue#16164](#), [pr#10104](#), Patrick Donnelly)
- mds: fix SnapRealm::have\_past\_parents\_open() ([issue#16299](#), [pr#10499](#), Yan, Zheng)
- mds: fix setattr starve setattr ([issue#16154](#), [pr#9560](#), Yan, Zheng)
- mds: wrongly treat symlink inode as normal file/dir when symlink inode is stale on kcephfs ([issue#15702](#), [pr#9405](#), Zhi Zhang)
- mon: “mon metadata” fails when only one monitor exists ([issue#15866](#), [pr#10654](#), John Spray, Kefu Chai)
- mon: Monitor: validate prefix on handle\_command() ([issue#16297](#), [pr#10036](#), You Ji)
- mon: OSDMonitor: drop pg temps from not the current primary ([issue#16127](#), [pr#9998](#), Samuel Just)

- mon: prepare\_pgtemp needs to only update up\_thru if newer than the existing one ([issue#16185](#), [pr#10001](#), Samuel Just)
- msgr: AsyncConnection::lockmsg/async lockdep cycle: AsyncMessenger::lock, MDSDaemon::mds\_lock, AsyncConnection::lock ([issue#16237](#), [pr#10004](#), Haomai Wang)
- msgr: async messenger mon crash ([issue#16378](#), [issue#16418](#), [pr#9996](#), Haomai Wang)
- msgr: backports of all asyncmsgr fixes to jewel ([issue#15503](#), [issue#15372](#), [pr#9633](#), Yan Jun, Haomai Wang, Piotr Dałek)
- msgr: msg/async: connection race hang ([issue#15849](#), [pr#10003](#), Haomai Wang)
- osd: FileStore: umount hang because sync thread doesn't exit ([issue#15695](#), [pr#9105](#), Kefu Chai)
- osd: Fixes for list-inconsistent-\* ([issue#15766](#), [issue#16192](#), [issue#15719](#), [pr#9565](#), David Zafman)
- osd: New pools have bogus stuck inactive/unclean HEALTH\_ERR messages until they are first active and clean ([issue#14952](#), [pr#10007](#), Sage Weil)
- osd: OSD crash with Hammer to Jewel Upgrade: void FileStore::init\_temp\_collections() ([issue#16672](#), [pr#10561](#), David Zafman)
- osd: OSD failed to subscribe skipped osdmaps after ceph osd pause ([issue#17023](#), [pr#10804](#), Kefu Chai)
- osd: ObjectCacher split BufferHead read fix ([issue#16002](#), [pr#10074](#), Greg Farnum)
- osd: ReplicatedBackend doesn't increment stats on pull, only push ([issue#16277](#), [pr#10421](#), Kefu Chai)
- osd: Scrub error: 0/1 pinned ([issue#15952](#), [pr#9576](#), Samuel Just)
- osd: crash adding snap to purged\_snaps in ReplicatedPG::WaitingOnReplicas ([issue#15943](#), [pr#9575](#), Samuel Just)
- osd: partprobe intermittent issues during ceph-disk prepare ([issue#15176](#), [pr#10497](#), Marius Vollmer, Loic Dachary)
- osd: saw valgrind issues in ReplicatedPG::new\_repop ([issue#16801](#), [pr#10760](#), Kefu Chai)
- osd: sparse\_read on ec pool should return extends with correct offset ([issue#16138](#), [pr#10006](#), kofiliu)
- osd:sched\_time not actually randomized ([issue#15890](#), [pr#9578](#), xie xingguo)
- rbd: ImageReplayer::is\_replaying does not include flush state ([issue#16970](#), [pr#10790](#), Jason Dillaman)

- rbd: Journal duplicate op detection can cause lockdep error ([issue#16363](#), [pr#10044](#), Jason Dillaman)
- rbd: Journal needs to handle duplicate maintenance op tids ([issue#16362](#), [pr#10045](#), Jason Dillaman)
- rbd: Unable to disable journaling feature if in unexpected mirror state ([issue#16348](#), [pr#10042](#), Jason Dillaman)
- rbd: bashism in src/rbdmap ([issue#16608](#), [pr#10786](#), Jason Dillaman)
- rbd: doc: format 2 now is the default image format ([issue#17026](#), [pr#10732](#), Chengwei Yang)
- rbd: hen journaling is enabled, a flush request shouldn't flush the cache ([issue#15761](#), [pr#10041](#), Yuan Zhou)
- rbd: possible race condition during journal transition from replay to ready ([issue#16198](#), [pr#10047](#), Jason Dillaman)
- rbd: qa/workunits/rbd: respect RBD\_CREATE\_ARGS environment variable ([issue#16289](#), [pr#9721](#), Mykola Golub)
- rbd: rbd-mirror should disable proxied maintenance ops for non-primary image ([issue#16411](#), [pr#10050](#), Jason Dillaman)
- rbd: rbd-mirror: FAILED assert(m\_local\_image\_ctx->object\_map != nullptr) ([issue#16558](#), [pr#10646](#), Jason Dillaman)
- rbd: rbd-mirror: FAILED assert(m\_on\_update\_status\_finish == nullptr) ([issue#16956](#), [pr#10792](#), Jason Dillaman)
- rbd: rbd-mirror: FAILED assert(m\_state == STATE\_STOPPING) ([issue#16980](#), [pr#10791](#), Jason Dillaman)
- rbd: rbd-mirror: ensure replay status formatter has completed before stopping replay ([issue#16352](#), [pr#10043](#), Jason Dillaman)
- rbd: rbd-mirror: include local pool id in resync throttle unique key ([issue#16536](#), [issue#15239](#), [issue#16488](#), [issue#16491](#), [issue#16329](#), [issue#15108](#), [issue#15670](#), [pr#10678](#), Ricardo Dias, Jason Dillaman)
- rbd: rbd-mirror: potential race condition accessing local image journal ([issue#16230](#), [pr#10046](#), Jason Dillaman)
- rbd: rbd-mirror: reduce memory footprint during journal replay ([issue#16321](#), [issue#16489](#), [issue#16622](#), [issue#16539](#), [issue#16223](#), [issue#16349](#), [pr#10684](#), Mykola Golub, Jason Dillaman)
- rgw: A query on a static large object fails with 404 error ([issue#16015](#), [pr#9544](#), Radoslaw Zarzynski)

- rgw: Add zone rename to radosgw\_admin ([issue#16934](#), [pr#10663](#), Shilpa Jagannath)
- rgw: Bucket index shards orphaned after bucket delete ([issue#16412](#), [pr#10525](#), Orit Wasserman)
- rgw: Bug when using port 443s in rgw. ([issue#16548](#), [pr#10664](#), Pritha Srivastava)
- rgw: Fallback to Host header for bucket name. ([issue#15975](#), [pr#10693](#), Robin H. Johnson)
- rgw: Fix civetweb IPv6 ([issue#16928](#), [pr#10580](#), Robin H. Johnson)
- rgw: Increase log level for messages occurring while running rgw admin command ([issue#16935](#), [pr#10765](#), Shilpa Jagannath)
- rgw: No Last-Modified, Content-Size and X-Object-Manifest headers if no segments in DLO manifest ([issue#15812](#), [pr#9265](#), Radoslaw Zarzynski)
- rgw: RGWPeriodPuller tries to pull from itself ([issue#16939](#), [pr#10764](#), Casey Bodley)
- rgw: Set Access-Control-Allow-Origin to a Asterisk if allowed in a rule ([issue#15348](#), [pr#9453](#), Wido den Hollander)
- rgw: Swift API returns double space usage and objects of account metadata ([issue#16188](#), [pr#10148](#), Albert Tu)
- rgw: account/container metadata not actually present in a request are deleted during POST through Swift API ([issue#15977](#), [issue#15779](#), [pr#9542](#), Radoslaw Zarzynski)
- rgw: add socket backlog setting for via ceph.conf ([issue#16406](#), [pr#10216](#), Feng Guo)
- rgw: add tenant support to multisite sync ([issue#16469](#), [issue#16121](#), [issue#16665](#), [pr#10845](#), Yehuda Sadeh, Josh Durgin, Casey Bodley, Pritha Srivastava)
- rgw: add\_zone only clears master\_zone if -master=false ([issue#15901](#), [pr#9327](#), Casey Bodley)
- rgw: aws4 parsing issue ([issue#15940](#), [issue#15939](#), [pr#9545](#), Yehuda Sadeh)
- rgw: aws4: add STREAMING-AWS4-HMAC-SHA256-PAYLOAD support ([issue#16146](#), [pr#10167](#), Radoslaw Zarzynski, Javier M. Mellid)
- rgw: backport merge of static sites fixes ([issue#15555](#), [issue#15532](#), [issue#15531](#), [pr#9568](#), Robin H. Johnson)
- rgw: can set negative max\_buckets on RGWUserInfo ([issue#14534](#), [pr#10655](#), Yehuda Sadeh)
- rgw: cleanup radosgw-admin temp command as it was deprecated ([issue#16023](#),

pr#9390, Vikhyat Umrao)

- rgw: comparing return code to ERR\_NOT\_MODIFIED in rgw\_rest\_s3.cc (needs minus sign) ([issue#16327](#), [pr#9790](#), Nathan Cutler)
- rgw: custom metadata aren't camelcased in Swift's responses ([issue#15902](#), [pr#9267](#), Radoslaw Zarzynski)
- rgw: data sync stops after getting error in all data log sync shards ([issue#16530](#), [pr#10073](#), Yehuda Sadeh)
- rgw: default zone and zonegroup cannot be added to a realm ([issue#16839](#), [pr#10658](#), Casey Bodley)
- rgw: document multi tenancy ([issue#16635](#), [pr#10217](#), Pete Zaitcev)
- rgw: don't unregister request if request is not connected to manager ([issue#15911](#), [pr#9242](#), Yehuda Sadeh)
- rgw: failed to create bucket after upgrade from hammer to jewel ([issue#16627](#), [pr#10524](#), Orit Wasserman)
- rgw: fix ldap bindpw parsing ([issue#16286](#), [pr#10518](#), Matt Benjamin)
- rgw: fix multi-delete query param parsing. ([issue#16618](#), [pr#10188](#), Robin H. Johnson)
- rgw: improve support for Swift's object versioning. ([issue#15925](#), [pr#10710](#), Radoslaw Zarzynski)
- rgw: initial slashes are not properly handled in Swift's BulkDelete ([issue#15948](#), [pr#9316](#), Radoslaw Zarzynski)
- rgw: master: build failures with boost > 1.58 ([issue#16392](#), [issue#16391](#), [pr#10026](#), Abhishek Lekshmanan)
- rgw: multisite segfault on ~RGWRealmWatcher if realm was deleted ([issue#16817](#), [pr#10660](#), Casey Bodley)
- rgw: multisite sync races with deletes ([issue#16222](#), [issue#16464](#), [issue#16220](#), [issue#16143](#), [pr#10293](#), Yehuda Sadeh, Casey Bodley)
- rgw: multisite: preserve zone's extra pool ([issue#16712](#), [pr#10537](#), Abhishek Lekshmanan)
- rgw: object expirer's hints might be trimmed without processing in some circumstances ([issue#16705](#), [issue#16684](#), [pr#10763](#), Radoslaw Zarzynski)
- rgw: radosgw-admin failure for user create after upgrade from hammer to jewel ([issue#15937](#), [pr#9294](#), Orit Wasserman, Abhishek Lekshmanan)
- rgw: radosgw-admin: EEXIST messages for create operations ([issue#15720](#), [pr#9268](#),

Abhishek Lekshmanan)

- rgw: radosgw-admin: inconsistency in uid/email handling ([issue#13598](#), [pr#10520](#), Matt Benjamin)
- rgw: realm pull fails when using apache frontend ([issue#15846](#), [pr#9266](#), Orit Wasserman)
- rgw: retry on bucket sync errors ([issue#16108](#), [pr#9425](#), Yehuda Sadeh)
- rgw: s3website: x-amz-website-redirect-location header returns malformed HTTP response ([issue#15531](#), [pr#9099](#), Robin H. Johnson)
- rgw: segfault in RGWOp\_MDLog\_Notify ([issue#16666](#), [pr#10662](#), Casey Bodley)
- rgw: segmentation fault on error\_repo in data sync ([issue#16603](#), [pr#10523](#), Casey Bodley)
- rgw: selinux denials in RGW ([issue#16126](#), [pr#10519](#), Boris Ranto)
- rgw: support size suffixes for -max-size in radosgw-admin command ([issue#16004](#), [pr#9743](#), Vikhyat Umrao)
- rgw: updating CORS/ACLs might not work in some circumstances ([issue#15976](#), [pr#9543](#), Radoslaw Zarzynski)
- rgw: use zone endpoints instead of zonegroup endpoints ([issue#16834](#), [pr#10659](#), Casey Bodley)
- tests: improve rbd-mirror test case coverage ([issue#16197](#), [pr#9631](#), Mykola Golub, Jason Dillaman)
- tests: rados/test.sh workunit timeout on OpenStack ([issue#15403](#), [pr#8904](#), Loic Dachary)
- tools: ceph-disk: Accept bcache devices as data disks ([issue#13278](#), [pr#8497](#), Peter Sabaini)
- tools: rados: Add cleanup message with time to rados bench output ([issue#15704](#), [pr#9740](#), Vikhyat Umrao)
- tools: src/script/subman fails with KeyError: 'nband' ([issue#16961](#), [pr#10625](#), Loic Dachary, Ali Maredia)

## v10.2.2 Jewel

---

This point release fixes several important bugs in RBD mirroring, RGW multi-site, CephFS, and RADOS.

We recommend that all v10.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

- ceph: cli: exception when pool name has non-ascii characters ([issue#15913](#), [pr#9320](#), Ricardo Dias)
- ceph-disk: workaround gperftool hang ([issue#13522](#), [issue#16103](#), [pr#9427](#), Loic Dachary)
- cephfs: backports needed for Manila ([issue#15599](#), [issue#15417](#), [issue#15045](#), [pr#9430](#), John Spray, Ramana Raja, Xiaoxi Chen)
- ceph.spec.in: drop support for RHEL<7 and SUSE<1210 in jewel and above ([issue#15725](#), [issue#15627](#), [issue#13445](#), [issue#15822](#), [issue#15472](#), [issue#15987](#), [issue#15516](#), [issue#15549](#), [pr#8938](#), Boris Ranto, Sage Weil, Nathan Cutler, Lars Marowsky-Bree)
- ceph\_test\_librbd\_fsx crashes during journal replay shut down ([issue#16123](#), [pr#9556](#), Jason Dillaman)
- client: fix bugs accidentally disabling readahead ([issue#16024](#), [pr#9656](#), Patrick Donnelly, Greg Farnum)
- cls\_journal: initialize empty commit position upon client register ([issue#15757](#), [pr#9376](#), runsi, Venky Shankar)
- cls::rbd: mirror\_image\_status\_list returned max 64 items ([pr#9069](#), Mykola Golub)
- cls\_rbd: mirror image status summary should read full directory ([issue#16178](#), [pr#9608](#), Jason Dillaman)
- common: BackoffThrottle spins unnecessarily with very small backoff while the throttle is full ([issue#15953](#), [pr#9579](#), Samuel Just)
- common: Do not link lttng into libglobal ([pr#9194](#), Karol Mroz)
- debian: install systemd target files ([issue#15573](#), [pr#8815](#), Kefu Chai, Sage Weil)
- doc: update mirroring guide to include pool/image status commands ([issue#15746](#), [pr#9180](#), Mykola Golub)
- librbd: Disabling journaling feature results in "Transport endpoint is not connected" error ([issue#15863](#), [pr#9548](#), Yuan Zhou)
- librbd: do not shut down exclusive lock while acquiring' ([issue#16291](#), [issue#16260](#), [pr#9691](#), Jason Dillaman)
- librbd: Initial python APIs to support mirroring ([issue#15656](#), [pr#9550](#), Mykola Golub)

- librbd: journal IO error results in failed assertion in AioCompletion ([issue#16077](#), [issue#15034](#), [issue#15791](#), [pr#9611](#), Hector Martin, Jason Dillaman)
- librbd: journal: live replay might skip entries from previous object set ([issue#15864](#), [issue#15665](#), [pr#9217](#), Jason Dillaman)
- librbd: journal: support asynchronous shutdown ([issue#15949](#), [issue#14530](#), [issue#15993](#), [pr#9373](#), Jason Dillaman)
- librbd: Metadata config overrides are applied synchronously ([issue#15928](#), [pr#9318](#), Jason Dillaman)
- librbd: Object Map is showing as invalid, even when Object Map is disabled for that Image. ([issue#16076](#), [pr#9555](#), xinxin shu)
- librbd: prevent error messages when journal externally disabled ([issue#16114](#), [pr#9610](#), Zhiqiang Wang, Jason Dillaman)
- librbd: recursive lock possible when disabling journaling ([issue#16235](#), [pr#9654](#), Jason Dillaman)
- librbd: refresh image if needed in mirror functions ([issue#16096](#), [pr#9609](#), Jon Bernard)
- librbd: remove should ignore mirror errors from older OSDs ([issue#16268](#), [pr#9692](#), Jason Dillaman)
- librbd: reuse ImageCtx::finisher and SafeTimer for lots of images case ([issue#13938](#), [pr#9580](#), Haomai Wang)
- librbd: validate image metadata configuration overrides ([issue#15522](#), [pr#9554](#), zhuangzeqiang)
- mds: order directories by hash and fix simultaneous readdir races ([issue#15508](#), [pr#9655](#), Yan, Zheng, Greg Farnum)
- mon: Hammer (0.94.3) OSD does not delete old OSD Maps in a timely fashion (maybe at all?) ([issue#13990](#), [pr#9100](#), Kefu Chai)
- mon/Monitor: memory leak on Monitor::handle\_ping() ([issue#15793](#), [pr#9270](#), xie xingguo)
- osd: acting\_primary not updated on split ([issue#15523](#), [pr#8968](#), Sage Weil)
- osd: boot race with noup being set ([issue#15678](#), [pr#9101](#), Sage Weil)
- osd: deadlock in OSD::\_committed\_osd\_maps ([issue#15701](#), [pr#9103](#), Xinze Chi)
- osd: hobject\_t::get\_max() vs is\_max() discrepancy ([issue#16113](#), [pr#9614](#), Samuel Just)
- osd: LibRadosWatchNotifyPPTests/LibRadosWatchNotifyPP.WatchNotify2Timeout/1 segv

([issue#15760](#), [pr#9104](#), Sage Weil)

- osd: remove reliance on FLAG\_OMAP for reads ([pr#9638](#), Samuel Just)
- osd valgrind invalid reads/writes ([issue#15870](#), [pr#9237](#), Samuel Just)
- pybind: rbd API should default features parameter to None ([issue#15982](#), [pr#9553](#), Mykola Golub)
- qa: dynamic\_features.sh races with image deletion ([issue#15500](#), [pr#9552](#), Mykola Golub)
- qa/workunits: ensure replay has started before checking position ([issue#16248](#), [pr#9674](#), Jason Dillaman)
- qa/workunits/rbd: fixed rbd\_mirror teuthology runtime errors ([pr#9232](#), Jason Dillaman)
- radosgw-admin: fix 'period push' handling of -url ([issue#15926](#), [pr#9210](#), Casey Bodley)
- rbd-mirror: Delete local image mirror when remote image mirroring is disabled ([issue#15916](#), [issue#14421](#), [pr#9372](#), runsisi, Mykola Golub, Ricardo Dias)
- rbd-mirror: do not propagate deletions when pool unavailable ([issue#16229](#), [pr#9630](#), Jason Dillaman)
- rbd-mirror: do not re-use image id from mirror directory if creating image ([issue#16253](#), [pr#9673](#), Jason Dillaman)
- rbd-mirror: FAILED assert(!m\_status\_watcher) ([issue#16245](#), [issue#16290](#), [pr#9690](#), Mykola Golub)
- rbd-mirror: fix deletion propagation edge cases ([issue#16226](#), [pr#9629](#), Jason Dillaman)
- rbd-mirror: fix journal shut down ordering ([issue#16165](#), [pr#9628](#), Jason Dillaman)
- rbd-mirror: potential crash during image status update ([issue#15909](#), [pr#9226](#), Mykola Golub, Jason Dillaman)
- rbd-mirror: refresh image after creating sync point ([issue#16196](#), [pr#9627](#), Jason Dillaman)
- rbd-mirror: replicate cloned images ([issue#14937](#), [pr#9423](#), Jason Dillaman)
- rbd-mirror should disable the rbd cache for local images ([issue#15930](#), [pr#9317](#), Jason Dillaman)
- rbd-mirror: support bootstrap canceling ([issue#16201](#), [pr#9612](#), Mykola Golub)
- rbd-mirror: support multiple replicated pools ([issue#16045](#), [pr#9409](#), Jason

Dillaman)

- rgw: fix manager selection when APIs customized ([issue#15974](#), [issue#15973](#), [pr#9245](#), Robin H. Johnson)
- rgw: keep track of written\_objs correctly ([issue#15886](#), [pr#9239](#), Yehuda Sadeh)
- rpm: ceph gid mismatch on upgrade from hammer with pre-existing ceph user (SUSE) ([issue#15869](#), [pr#9424](#), Nathan Cutler)
- systemd: ceph-{mds,mon,osd,radosgw} systemd unit files need wants=time-sync.target ([issue#15419](#), [pr#8802](#), Nathan Cutler)
- test: failure in journal.sh workunit test ([issue#16011](#), [pr#9377](#), Mykola Golub)
- tests: rm -fr /tmp/virtualenv ([issue#16087](#), [pr#9403](#), Loic Dachary)

## v10.2.1 Jewel

---

This is the first bugfix release for Jewel. It contains several annoying packaging and init system fixes and a range of important bugfixes across RBD, RGW, and CephFS.

We recommend that all v10.2.x users upgrade.

For more detailed information, see [the complete changelog](#).

# Notable Changes

---

- cephfs: CephFSVolumeClient should isolate volumes by RADOS namespace ([issue#15400](#), [pr#8787](#), Xiaoxi Chen)
- cephfs: handle standby-replay nodes properly in upgrades ([issue#15591](#), [pr#8971](#), John Spray)
- ceph-{mds,mon,osd} packages need scriptlets with systemd code ([issue#14941](#), [pr#8801](#), Boris Ranto, Nathan Cutler)
- ceph\_test\_keyvaluedb: fix ([issue#15435](#), [pr#9051](#), Allen Samuels, Sage Weil)
- cmake: add missing source file to rbd\_mirror/image\_replayer ([pr#9052](#), Casey Bodley)
- cmake: fix rbd compile errors ([pr#9076](#), runsisi, Jason Dillaman)
- journal: incorrectly computed object offset within set ([issue#15765](#), [pr#9038](#), Jason Dillaman)
- librbd: client-side handling for incompatible object map sizes ([issue#15642](#), [pr#9039](#), Jason Dillaman)
- librbd: constrain size of AioWriteEvent journal entries ([issue#15750](#), [pr#9048](#), Jason Dillaman)
- librbd: does not crash if image header is too short ([pr#9044](#), Kefu Chai)
- librbd: Errors encountered disabling object-map while flatten is in-progress ([issue#15572](#), [pr#8869](#), Jason Dillaman)
- librbd: fix get/list mirror image status API ([issue#15771](#), [pr#9036](#), Mykola Golub)
- librbd: Parent image is closed twice if error encountered while opening ([issue#15574](#), [pr#8867](#), Jason Dillaman)
- librbd: possible double-free of object map invalidation request upon error ([issue#15643](#), [pr#8865](#), runsisi)
- librbd: possible race condition leads to use-after-free ([issue#15690](#), [pr#9009](#), Jason Dillaman)
- librbd: potential concurrent event processing during journal replay ([issue#15755](#), [pr#9040](#), Jason Dillaman)
- librbd: Potential double free of SetSnapRequest instance ([issue#15571](#), [pr#8803](#), runsisi)
- librbd: put the validation of image snap context earlier ([pr#9046](#), runsisi)

- librbd: reduce log level for image format 1 warning ([issue#15577](#), [pr#9003](#), Jason Dillaman)
- mds/MDSAuthCap parse no longer fails on paths with hyphens ([issue#15465](#), [pr#8969](#), John Spray)
- mds: MDS incarnation no longer gets lost after remove filesystem ([issue#15399](#), [pr#8970](#), John Spray)
- mon/OSDMonitor: avoid underflow in reweight-by-utilization if max\_change=1 ([issue#15655](#), [pr#9006](#), Samuel Just)
- python: clone operation will fail if config overridden with "rbd default format = 1" ([issue#15685](#), [pr#8972](#), Jason Dillaman)
- radosgw-admin: add missing -zonegroup-id to usage ([issue#15650](#), [pr#9019](#), Casey Bodley)
- radosgw-admin: update usage for zone[group] modify ([issue#15651](#), [pr#9016](#), Casey Bodley)
- radosgw-admin: zonegroup remove command ([issue#15684](#), [pr#9015](#), Casey Bodley)
- rbd CLI to retrieve rbd mirror state for a pool / specific image ([issue#15144](#), [issue#14420](#), [pr#8868](#), Mykola Golub)
- rbd disk-usage CLI command should support calculating full image usage ([issue#14540](#), [pr#8870](#), Jason Dillaman)
- rbd: helpful error message on map failure ([issue#15721](#), [pr#9041](#), Venky Shankar)
- rbd: help message distinction between commands and aliases ([issue#15521](#), [pr#9004](#), Yongqiang He)
- rbd-mirror: admin socket commands to start/stop/restart mirroring ([issue#15718](#), [pr#9010](#), Mykola Golub, Josh Durgin)
- rbd-mirror can crash if start up is interrupted ([issue#15630](#), [pr#8866](#), Jason Dillaman)
- rbd-mirror: image sync needs to handle snapshot size and protection status ([issue#15110](#), [pr#9050](#), Jason Dillaman)
- rbd-mirror: lockdep error during bootstrap ([issue#15664](#), [pr#9008](#), Jason Dillaman)
- rbd-nbd: fix rbd-nbd aio callback error handling ([issue#15604](#), [pr#9005](#), Chang-Yi Lee)
- rgw: add AWS4 completion support for RGW\_OP\_SET\_BUCKET\_WEBSITE ([issue#15626](#), [pr#9018](#), Javier M. Mellid)
- rgw admin output ([issue#15747](#), [pr#9054](#), Casey Bodley)

- rgw: fix issue #15597 ([issue#15597](#), [pr#9020](#), Yehuda Sadeh)
- rgw: fix printing wrong X-Storage-Url in Swift's TempAuth. ([issue#15667](#), [pr#9021](#), Radoslaw Zarzynski)
- rgw: handle stripe transition when flushing final pending\_data.bl ([issue#15745](#), [pr#9053](#), Yehuda Sadeh)
- rgw: leak fixes ([issue#15792](#), [pr#9022](#), Yehuda Sadeh)
- rgw: multisite: Issues with Deleting Buckets ([issue#15540](#), [pr#8930](#), Abhishek Lekshmanan)
- rgw: period commit fix ([issue#15828](#), [pr#9081](#), Casey Bodley)
- rgw: period delete fixes ([issue#15469](#), [pr#9047](#), Casey Bodley)
- rgw: radosgw-admin zone set cuts pool names short if name starts with a period ([issue#15598](#), [pr#9029](#), Yehuda Sadeh)
- rgw: segfault at RGWAsyncGetSystemObj ([issue#15565](#), [issue#15625](#), [pr#9017](#), Yehuda Sadeh)
- several backports ([issue#15588](#), [issue#15655](#), [pr#8853](#), Alexandre Derumier, xie xingguo, Alfredo Deza)
- systemd: fix typo in preset file ([pr#8843](#), Nathan Cutler)
- tests: make check fails on ext4 ([issue#15837](#), [pr#9063](#), Loic Dachary, Sage Weil)

## v10.2.0 Jewel

---

This major release of Ceph is the foundation for the next long-term stable release series. There have been many major changes since the Infernalis (9.2.x) and Hammer (0.94.x) releases, and the upgrade process is non-trivial. Please read these release notes carefully.

## Major Changes from Infernalis

---

- *CephFS*:
  - This is the first release in which CephFS is declared stable! Several features are disabled by default, including snapshots and multiple active MDS servers.
  - The repair and disaster recovery tools are now feature-complete.
  - A new cephfs-volume-manager module is included that provides a high-level interface for creating “shares” for OpenStack Manila and similar projects.

- There is now experimental support for multiple CephFS file systems within a single cluster.
- *RGW:*
  - The multisite feature has been almost completely rearchitected and rewritten to support any number of clusters/sites, bidirectional fail-over, and active/active configurations.
  - You can now access radosgw buckets via NFS (experimental).
  - The AWS4 authentication protocol is now supported.
  - There is now support for S3 request payer buckets.
  - The new multitenancy infrastructure improves compatibility with Swift, which provides a separate container namespace for each user/tenant.
  - The OpenStack Keystone v3 API is now supported. There are a range of other small Swift API features and compatibility improvements as well, including bulk delete and SLO (static large objects).
- *RBD:*
  - There is new support for mirroring (asynchronous replication) of RBD images across clusters. This is implemented as a per-RBD image journal that can be streamed across a WAN to another site, and a new rbd-mirror daemon that performs the cross-cluster replication.
  - The exclusive-lock, object-map, fast-diff, and journaling features can be enabled or disabled dynamically. The deep-flatten features can be disabled dynamically but not re-enabled.
  - The RBD CLI has been rewritten to provide command-specific help and full bash completion support.
  - RBD snapshots can now be renamed.
- *RADOS:*
  - BlueStore, a new OSD backend, is included as an experimental feature. The plan is for it to become the default backend in the K or L release.
  - The OSD now persists scrub results and provides a librados API to query results in detail.
  - We have revised our documentation to recommend *against* using ext4 as the underlying filesystem for Ceph OSD daemons due to problems supporting our long object name handling.

# Major Changes from Hammer

- *General:*
  - Ceph daemons are now managed via systemd (with the exception of Ubuntu Trusty, which still uses upstart).
  - Ceph daemons run as ‘ceph’ user instead of ‘root’.
  - On Red Hat distros, there is also an SELinux policy.
- *RADOS:*
  - The RADOS cache tier can now proxy write operations to the base tier, allowing writes to be handled without forcing migration of an object into the cache.
  - The SHEC erasure coding support is no longer flagged as experimental. SHEC trades some additional storage space for faster repair.
  - There is now a unified queue (and thus prioritization) of client IO, recovery, scrubbing, and snapshot trimming.
  - There have been many improvements to low-level repair tooling (ceph-objectstore-tool).
  - The internal ObjectStore API has been significantly cleaned up in order to facilitate new storage backends like BlueStore.
- *RGW:*
  - The Swift API now supports object expiration.
  - There are many Swift API compatibility improvements.
- *RBD:*
  - The `rbd du` command shows actual usage (quickly, when object-map is enabled).
  - The object-map feature has seen many stability improvements.
  - The object-map and exclusive-lock features can be enabled or disabled dynamically.
  - You can now store user metadata and set persistent librbd options associated with individual images.
  - The new deep-flatten features allow flattening of a clone and all of its snapshots. (Previously snapshots could not be flattened.)

- The `export-diff` command is now faster (it uses aio). There is also a new `fast-diff` feature.
- The `-size` argument can be specified with a suffix for units (e.g., `--size 64G`).
- There is a new `rbd status` command that, for now, shows who has the image open/mapped.
- *CephFS*:
  - You can now rename snapshots.
  - There have been ongoing improvements around administration, diagnostics, and the check and repair tools.
  - The caching and revocation of client cache state due to unused inodes has been dramatically improved.
  - The `ceph-fuse` client behaves better on 32-bit hosts.

## Distro compatibility

---

Starting with Infernalis, we have dropped support for many older distributions so that we can move to a newer compiler toolchain (e.g., C++11). Although it is still possible to build Ceph on older distributions by installing backported development tools, we are not building and publishing release packages for ceph.com.

We now build packages for the following distributions and architectures:

- `x86_64`:
  - CentOS 7.x. We have dropped support for CentOS 6 (and other RHEL 6 derivatives, like Scientific Linux 6).
  - Debian Jessie 8.x. Debian Wheezy 7.x's g++ has incomplete support for C++11 (and no systemd).
  - Ubuntu Xenial 16.04 and Trusty 14.04. Ubuntu Precise 12.04 is no longer supported.
  - Fedora 22 or later.
- `aarch64 / arm64`:
  - Ubuntu Xenial 16.04.

# Upgrading from Infernalis or Hammer

- We now recommend against using `ext4` as the underlying file system for Ceph OSDs, especially when RGW or other users of long RADOS object names are used. For more information about why, please see [Filesystem Recommendations](#).

If you have an existing cluster that uses ext4 for the OSDs but uses only RBD and/or CephFS, then the ext4 limitations will not affect you. Before upgrading, be sure add the following to `ceph.conf` to allow the OSDs to start:

```
1. osd max object name len = 256
2. osd max object namespace len = 64
```

Keep in mind that if you set these lower object name limits and later decide to use RGW on this cluster, it will have problems storing S3/Swift objects with long names. This startup check can also be disabled via the below option, although this is not recommended:

```
1. osd check max object name len on startup = false
```

- There are no major compatibility changes since Infernalis. Simply upgrading the daemons on each host and restarting all daemons is sufficient.
- The rbd CLI no longer accepts the deprecated ‘-image-features’ option during create, import, and clone operations. The ‘-image-feature’ option should be used instead.
- The rbd legacy image format (version 1) is deprecated with the Jewel release. Attempting to create a new version 1 RBD image will result in a warning. Future releases of Ceph will remove support for version 1 RBD images.
- The ‘send\_pgCreates’ and ‘map\_pgCreates’ mon CLI commands are obsolete and no longer supported.
- A new configure option ‘mon\_election\_timeout’ is added to specifically limit max waiting time of monitor election process, which was previously restricted by ‘mon\_lease’.
- CephFS filesystems created using versions older than Firefly (0.80) must use the new ‘cephfs-data-scan tmap\_upgrade’ command after upgrading to Jewel. See ‘Upgrading’ in the CephFS documentation for more information.
- The ‘ceph mds setmap’ command has been removed.
- The default RBD image features for new images have been updated to enable the following: exclusive lock, object map, fast-diff, and deep-flatten. These features are not currently supported by the RBD kernel driver nor older RBD

clients. They can be disabled on a per-image basis via the RBD CLI, or the default features can be updated to the pre-Jewel setting by adding the following to the client section of the Ceph configuration file:

```
1. rbd default features = 1
```

- The rbd legacy image format (version 1) is deprecated with the Jewel release.
- After upgrading, users should set the ‘sortbitwise’ flag to enable the new internal object sort order:

```
1. ceph osd set sortbitwise
```

This flag is important for the new object enumeration API and for new backends like BlueStore.

- The rbd CLI no longer permits creating images and snapshots with potentially ambiguous names (e.g. the ‘/’ and ‘@’ characters are disallowed). The validation can be temporarily disabled by adding “–rbd-validate-names=false” to the rbd CLI when creating an image or snapshot. It can also be disabled by adding the following to the client section of the Ceph configuration file:

```
1. rbd validate names = false
```

## Upgrading from Hammer

- All cluster nodes must first upgrade to Hammer v0.94.4 or a later v0.94.z release; only then is it possible to upgrade to Jewel 10.2.z.
- For all distributions that support systemd (CentOS 7, Fedora, Debian Jessie 8.x, OpenSUSE), ceph daemons are now managed using native systemd files instead of the legacy sysvinit scripts. For example:

```
1. systemctl start ceph.target      # start all daemons
2. systemctl status ceph-osd@12     # check status of osd.12
```

The main notable distro that is *not* yet using systemd is Ubuntu trusty 14.04. (The next Ubuntu LTS, 16.04, will use systemd instead of upstart.)

- Ceph daemons now run as user and group `ceph` by default. The `ceph` user has a static UID assigned by Fedora and Debian (also used by derivative distributions like RHEL/CentOS and Ubuntu). On SUSE the same UID/GID as in Fedora and Debian will be used, *provided it is not already assigned*. In the unlikely event the preferred UID or GID is assigned to a different user/group, `ceph` will get a dynamically assigned UID/GID.

If your systems already have a ceph user, upgrading the package will cause problems. We suggest you first remove or rename the existing ‘ceph’ user and ‘ceph’ group before upgrading.

When upgrading, administrators have two options:

i. Add the following line to `ceph.conf` on all hosts:

```
1. setuser match path = /var/lib/ceph/$type/$cluster-$id
```

This will make the Ceph daemons run as root (i.e., not drop privileges and switch to user ceph) if the daemon’s data directory is still owned by root. Newly deployed daemons will be created with data owned by user ceph and will run with reduced privileges, but upgraded daemons will continue to run as root.

ii. Fix the data ownership during the upgrade. This is the preferred option, but it is more work and can be very time consuming. The process for each host is to:

a. Upgrade the ceph package. This creates the ceph user and group. For example:

```
1. ceph-deploy install --stable jewel HOST
```

b. Stop the daemon(s):

```
1. service ceph stop          # fedora, centos, rhel, debian
2. stop ceph-all              # ubuntu
```

c. Fix the ownership:

```
1. chown -R ceph:ceph /var/lib/ceph
2. chown -R ceph:ceph /var/log/ceph
```

d. Restart the daemon(s):

```
1. start ceph-all            # ubuntu
2. systemctl start ceph.target # debian, centos, fedora, rhel
```

*Alternatively, the same process can be done with a single daemon type, for example by stopping*

*1. only monitors and chowning only `/var/lib/ceph/mon`.*

- The on-disk format for the experimental KeyValueStore OSD backend has changed. You will need to remove any OSDs using that backend before you upgrade any test clusters that use it.
- When a pool quota is reached, librados operations now block indefinitely, the same way they do when the cluster fills up. (Previously they would return -ENOSPC.) By default, a full cluster or pool will now block. If your librados application can handle ENOSPC or EDQUOT errors gracefully, you can get error

returns instead by using the new librados OPERATION\_FULL\_TRY flag.

- The return code for librbd's rbd\_aio\_read and Image::aio\_read API methods no longer returns the number of bytes read upon success. Instead, it returns 0 upon success and a negative value upon failure.
- 'ceph scrub', 'ceph compact' and 'ceph sync force' are now DEPRECATED. Users should instead use 'ceph mon scrub', 'ceph mon compact' and 'ceph mon sync force'.
- 'ceph mon\_metadata' should now be used as 'ceph mon metadata'. There is no need to deprecate this command (same major release since it was first introduced).
- The -dump-json option of "osdmaptool" is replaced by -dump json.
- The commands of "pg ls-by-{pool,primary,osd}" and "pg ls" now take "recovering" instead of "recovery", to include the recovering pgs in the listed pgs.

# Upgrading from Firefly

Upgrading directly from Firefly v0.80.z is not recommended. It is possible to do a direct upgrade, but not without downtime, as all OSDs must be stopped, upgraded, and then restarted. We recommend that clusters be first upgraded to Hammer v0.94.6 or a later v0.94.z release; only then is it possible to upgrade to Jewel 10.2.z for an online upgrade (see below).

To do an offline upgrade directly from Firefly, all Firefly OSDs must be stopped and marked down before any Jewel OSDs will be allowed to start up. This fencing is enforced by the Jewel monitor, so you should use an upgrade procedure like:

1. Upgrade Ceph on monitor hosts
2. Restart all ceph-mon daemons
3. Set noout::  

```
ceph osd set noout
```
4. Upgrade Ceph on all OSD hosts
5. Stop all ceph-osd daemons
6. Mark all OSDs down with something like::  

```
ceph osd down seq 0 1000
```
7. Start all ceph-osd daemons
8. Let the cluster settle and then unset noout::  

```
ceph osd unset noout
```
9. Upgrade and restart any remaining daemons (ceph-mds, radosgw)

## Notable Changes since Infernalis

- aarch64: add optimized version of crc32c (Yazen Ghannam, Steve Capper)
- Adding documentation on how to use new dynamic throttle scheme ([pr#8069](#), Somnath Roy)
- admin/build-doc: depend on zlib1g-dev and graphviz ([pr#7522](#), Ken Dreyer)
- auth: cache/reuse crypto lib key objects, optimize msg signature check (Sage Weil)
- auth: fail if rotating key is missing (do not spam log) ([pr#6473](#), Qiankun Zheng)
- auth: fix crash when bad keyring is passed ([pr#6698](#), Dunrong Huang)

- auth: make keyring without mon entity type return -EACCES ([pr#5734](#), Xiaowei Chen)
- AUTHORS: update email ([pr#7854](#), Yehuda Sadeh)
- auth: reinit NSS after fork() (#11128 Yan, Zheng)
- authtool: update --help and manpage to match code. ([pr#8456](#), Robin H. Johnson)
- autotools: fix out of tree build (Krzysztof Kosinski)
- autotools: improve make check output (Loic Dachary)
- Be more careful about directory fragmentation and scrubbing ([issue#15167](#), [pr#8180](#), Yan, Zheng)
- bluestore: latest and greatest ([issue#14210](#), [issue#13801](#), [pr#6896](#), xie.xingguo, Jianpeng Ma, YiQiang Chen, Sage Weil, Ning Yao)
- buffer: add invalidate\_crc() (Piotr Dalek)
- buffer: add symmetry operator==() and operator!=() ([pr#7974](#), Kefu Chai)
- buffer: fix internal iterator invalidation on rebuild, get\_contiguous ([pr#6962](#), Sage Weil)
- buffer: fix zero bug (#12252 Haomai Wang)
- buffer: hide iterator\_impl symbols ([issue#14788](#), [pr#7688](#), Kefu Chai)
- buffer: increment history alloc as well in raw\_combined ([issue#14955](#), [pr#7910](#), Samuel Just)
- buffer: make usable outside of ceph source again ([pr#6863](#), Josh Durgin)
- buffer: raw\_combined allocations buffer and ref count together ([pr#7612](#), Sage Weil)
- buffer: some cleanup (Michal Jarzabek)
- buffer: use move construct to append/push\_back/push\_front ([pr#7455](#), Haomai Wang)
- build: Adding build requires ([pr#7742](#), Erwan Velu)
- build: a few armhf (32-bit build) fixes ([pr#7999](#), Eric Lee, Sage Weil)
- build: allow jemalloc with rocksdb-static ([pr#7368](#), Somnath Roy)
- build: allow tcmalloc-minimal (Thorsten Behrens)
- build: build internal plugins and classes as modules ([pr#6462](#), James Page)
- build: C++11 now supported

- build: cmake check fixes ([pr#6787](#), Orit Wasserman)
- build: cmake: fix nss linking (Danny Al-Gaaf)
- build: cmake: misc fixes (Orit Wasserman, Casey Bodley)
- build: cmake tweaks ([pr#6254](#), John Spray)
- build: disable LTTNG by default (#11333 Josh Durgin)
- build: do not build ceph-dencoder with tcmalloc (#10691 Boris Ranto)
- build: fix a few warnings ([pr#6847](#), Orit Wasserman)
- build: fix bz2-dev dependency ([pr#6948](#), Samuel Just)
- build: fix compiling warnings ([pr#8366](#), Dongsheng Yang)
- build: Fixing BTRFS issue at 'make check' ([pr#7805](#), Erwan Velu)
- build: fix Jenkins make check errors due to deep-scrub randomization ([pr#6671](#), David Zafman)
- build: fix junit detection on Fedora 22 (Ira Cooper)
- build: fix pg ref disabling (William A. Kennington III)
- build: fix ppc build (James Page)
- build: fix the autotools and cmake build (the new fusestore needs libfuse) ([pr#7393](#), Kefu Chai)
- build: fix warnings ([pr#7197](#), Kefu Chai, xie xingguo)
- build: fix warnings ([pr#7315](#), Kefu Chai)
- build: FreeBSD related fixes ([pr#7170](#), Mykola Golub)
- build: Gentoo: \_FORTIFY\_SOURCE fix. ([issue#13920](#), [pr#6739](#), Robin H. Johnson)
- build: install-deps: misc fixes (Loic Dachary)
- build: install-deps.sh improvements (Loic Dachary)
- build: install-deps: support OpenSUSE (Loic Dachary)
- build: kill warnings ([pr#7397](#), Kefu Chai)
- build: make\_dist\_tarball.sh (Sage Weil)
- build: many cmake improvements
- build: misc cmake fixes (Matt Benjamin)

- build: misc fixes (Boris Ranto, Ken Dreyer, Owen Synge)
- build: misc make check fixes ([pr#7153](#), Sage Weil)
- build: more CMake package check fixes ([pr#6108](#), Daniel Gryniewicz)
- build: move libexec scripts to standardize across distros ([issue#14687](#), [issue#14705](#), [issue#14723](#), [pr#7636](#), Nathan Cutler, Kefu Chai)
- build/ops: enable CR in CentOS 7 ([issue#13997](#), [pr#6844](#), Loic Dachary)
- build/ops: rbd-replay moved from ceph-test-dbg to ceph-common-dbg ([issue#13785](#), [pr#6578](#), Loic Dachary)
- build/ops: systemd ceph-disk unit must not assume /bin/flock ([issue#13975](#), [pr#6803](#), Loic Dachary)
- build: OSX build fixes (Yan, Zheng)
- build: Refrain from versioning and packaging EC testing plugins ([issue#14756](#), [issue#14723](#), [pr#7637](#), Nathan Cutler, Kefu Chai)
- build: remove rest-bench
- build: Respect TMPDIR for virtualenv. ([pr#8457](#), Robin H. Johnson)
- build: spdk submodule; cmake ([pr#7503](#), Kefu Chai)
- build: workaround an automake bug for "make check" ([issue#14723](#), [pr#7626](#), Kefu Chai)
- ceph-authtool: fix return code on error (Gerhard Muntingh)
- ceph: bash auto complete for CLI based on mon command descriptions ([pr#7693](#), Adam Kupczyk)
- ceph\_daemon.py: Resolved ImportError to work with python3 ([pr#7937](#), Sarthak Munshi)
- ceph-detect-init: add debian/jessie test ([pr#8074](#), Kefu Chai)
- ceph-detect-init: added Linux Mint (Michal Jarzabek)
- ceph-detect-init: add missing test case ([pr#8105](#), Nathan Cutler)
- ceph-detect-init: fix py3 test ([pr#7025](#), Kefu Chai)
- ceph-detect-init: fix py3 test ([pr#7243](#), Kefu Chai)
- ceph\_detect\_init/\_\_init\_\_.py: remove shebang ([pr#7731](#), Nathan Cutler)
- ceph-detect-init: return correct value on recent SUSE distros ([issue#14770](#), [pr#7909](#), Nathan Cutler)

- ceph-detect-init: robust init system detection (Owen Synge)
- ceph-detect-init/run-tox.sh: FreeBSD: No init detect ([pr#8373](#), Willem Jan Withagen)
- ceph-detect-init: Ubuntu >= 15.04 uses systemd ([pr#6873](#), James Page)
- ceph-disk: Add destroy and deactivate option ([issue#7454](#), [pr#5867](#), Vicente Cheng)
- ceph-disk: add -f flag for btrfs mkfs ([pr#7222](#), Darrell Enns)
- ceph-disk: Add -setuser and -setgroup options for ceph-disk ([pr#7351](#), Mike Shuey)
- ceph-disk: ceph-disk list fails on /dev/cciss!c0d0 ([issue#13970](#), [issue#14233](#), [issue#14230](#), [pr#6879](#), Loic Dachary)
- ceph-disk: compare parted output with the dereferenced path ([issue#13438](#), [pr#6219](#), Joe Julian)
- ceph-disk: deactivate / destroy PATH arg are optional ([pr#7756](#), Loic Dachary)
- ceph-disk: do not always fail when re-using a partition ([pr#8508](#), You Ji)
- ceph-disk: ensure 'zap' only operates on a full disk (#11272 Loic Dachary)
- ceph-disk: fixes to respect init system (Loic Dachary, Owen Synge)
- ceph-disk: fix failures when preparing disks with udev > 214 ([issue#14080](#), [issue#14094](#), [pr#6926](#), Loic Dachary, Ilya Dryomov)
- ceph-disk: fix prepare -help ([pr#7758](#), Loic Dachary)
- ceph-disk: Fix trivial typo ([pr#7472](#), Brad Hubbard)
- ceph-disk: fix zap sgdisk invocation (Owen Synge, Thorsten Behrens)
- ceph-disk: flake8 fixes ([pr#7646](#), Loic Dachary)
- ceph-disk: follow ceph-osd hints when creating journal (#9580 Sage Weil)
- ceph-disk: get Nonetype when ceph-disk list with -format plain on single device. ([pr#6410](#), Vicente Cheng)
- ceph-disk: handle re-using existing partition (#10987 Loic Dachary)
- ceph-disk: improve parted output parsing (#10983 Loic Dachary)
- ceph-disk: Improving 'make check' for ceph-disk ([pr#7762](#), Erwan Velu)
- ceph-disk: install pip > 6.1 (#11952 Loic Dachary)
- ceph-disk: key management support ([issue#14669](#), [pr#7552](#), Loic Dachary)

- ceph-disk: make some arguments as required if necessary ([pr#7687](#), Dongsheng Yang)
- ceph-disk: make suppression work for activate-all and activate-journal (Dan van der Ster)
- ceph-disk: many fixes (Loic Dachary, Alfredo Deza)
- ceph-disk: pass -cluster arg on prepare subcommand (Kefu Chai)
- ceph-disk: s/dmcrpyt/dmcrypt/ ([issue#14838](#), [pr#7744](#), Loic Dachary, Frode Sandholtbraaten)
- ceph-disk: support bluestore ([issue#13422](#), [pr#7218](#), Loic Dachary, Sage Weil)
- ceph-disk: support for multipath devices (Loic Dachary)
- ceph-disk: support NVMe device partitions (#11612 Ilja Slepnev)
- ceph-disk/test: fix test\_prepare.py::TestPrepare tests ([pr#7549](#), Kefu Chai)
- ceph-disk: warn for prepare partitions with bad GUIDs ([issue#13943](#), [pr#6760](#), David Disseldorf)
- ceph: fix 'df' units (Zhe Zhang)
- ceph: fix parsing in interactive cli mode (#11279 Kefu Chai)
- ceph: fix tell behavior ([pr#6329](#), David Zafman)
- cephfs-data-scan: many additions, improvements (John Spray)
- cephfs-data-scan: scan\_frags ([pr#5941](#), John Spray)
- cephfs-data-scan: scrub tag filtering (#12133 and #12145) ([issue#12133](#), [issue#12145](#), [pr#5685](#), John Spray)
- ceph-fuse: add process to ceph-fuse -help ([pr#6821](#), Wei Feng)
- ceph-fuse: do not require successful remount when unmounting (#10982 Greg Farnum)
- ceph-fuse: fix double decreasing the count to trim caps ([issue#14319](#), [pr#7229](#), Zhi Zhang)
- ceph-fuse: fix double free of args ([pr#7015](#), Ilya Shipitsin)
- ceph-fuse: fix fsync() ([pr#6388](#), Yan, Zheng)
- ceph-fuse: Fix potential filehandle ref leak at umount ([issue#14800](#), [pr#7686](#), Zhi Zhang)
- ceph-fuse, libcephfs: don't clear COMPLETE when trimming null (Yan, Zheng)
- ceph-fuse, libcephfs: drop inode when rmdir finishes (#11339 Yan, Zheng)

- ceph-fuse,libcephfs: Fix client handling of “lost” open directories on shutdown ([issue#14996](#), [pr#7994](#), Yan, Zheng)
- ceph-fuse,libcephfs: fix free fds being exhausted eventually because freed fds are never put back ([issue#14798](#), [pr#7685](#), Zhi Zhang)
- ceph-fuse,libcephfs: fix uninlne (#11356 Yan, Zheng)
- ceph-fuse, libcephfs: hold exclusive caps on dirs we “own” (#11226 Greg Farnum)
- ceph-fuse: mostly behave on 32-bit hosts (Yan, Zheng)
- ceph-fuse:print usage information when no parameter specified ([pr#6868](#), Bo Cai)
- ceph-fuse: rotate log file ([pr#8485](#), Sage Weil)
- ceph-fuse: While starting ceph-fuse, start the log thread first ([issue#13443](#), [pr#6224](#), Wenjun Huang)
- ceph: improve error output for ‘tell’ (#11101 Kefu Chai)
- ceph: improve the error message ([issue#11101](#), [pr#7106](#), Kefu Chai)
- ceph.in: avoid a broken pipe error when use ceph command ([issue#14354](#), [pr#7212](#), Bo Cai)
- ceph.in: correct dev python path for automake builds ([pr#8360](#), Josh Durgin)
- ceph.in: fix python libpath for automake as well ([pr#8362](#), Josh Durgin)
- ceph.in: Minor python3 specific changes ([pr#7947](#), Sarthak Munshi)
- ceph-kvstore-tool: handle bad out file on command line ([pr#6093](#), Kefu Chai)
- ceph-mds:add -help/-h ([pr#6850](#), Cilang Zhao)
- ceph-monstore-tool: fix store-copy (Huangjun)
- ceph: new ‘ceph daemonperf’ command (John Spray, Mykola Golub)
- ceph\_objectstore\_bench: fix race condition, bugs ([issue#13516](#), [pr#6681](#), Igor Fedotov)
- ceph-objectstore-tool: fix -dry-run for many ceph-objectstore-tool operations ([pr#6545](#), David Zafman)
- ceph-objectstore-tool: many many improvements (David Zafman)
- ceph-objectstore-tool: refactoring and cleanup (John Spray)
- ceph-post-file: misc fixes (Joey McDonald, Sage Weil)
- ceph-rest-api: fix fs/flag/set ([pr#8428](#), Sage Weil)

- ceph.spec.in: add BuildRequires: systemd ([issue#13860](#), [pr#6692](#), Nathan Cutler)
- ceph.spec.in: add copyright notice ([issue#14694](#), [pr#7569](#), Nathan Cutler)
- ceph.spec.in: add license declaration ([pr#7574](#), Nathan Cutler)
- ceph.spec.in: disable lttng and babeltrace explicitly ([issue#14844](#), [pr#7857](#), Kefu Chai)
- ceph.spec.in: do not install Ceph RA on systemd platforms ([issue#14828](#), [pr#7894](#), Nathan Cutler)
- ceph.spec.in: fix openldap and openssl build dependencies for SUSE ([issue#15138](#), [pr#8120](#), Nathan Cutler)
- ceph.spec.in: limit \_smp\_mflags when lowmem\_builder is set in SUSE's OBS ([issue#13858](#), [pr#6691](#), Nathan Cutler)
- ceph\_test\_libcephfs: tolerate duplicated entries in readdir ([issue#14377](#), [pr#7246](#), Yan, Zheng)
- ceph\_test\_msgr: reduce test size to fix memory size ([pr#8127](#), Haomai Wang)
- ceph\_test\_msgr: Use send\_message instead of keepalive to wakeup connection ([pr#6605](#), Haomai Wang)
- ceph\_test\_rados\_misc: shorten mount timeout ([pr#8209](#), Sage Weil)
- ceph\_test\_rados: test pipelined reads (Zhiqiang Wang)
- check-generated.sh: can't source bash from sh ([pr#8521](#), Michal Jarzabek)
- cleanup ([pr#8058](#), Yehuda Sadeh, Orit Wasserman)
- cleanup: remove misc dead code ([pr#7201](#), Erwan Velu)
- client: a better check for MDS availability ([pr#6253](#), John Spray)
- client: add option to control how directory size is calculated ([pr#7323](#), Yan, Zheng)
- client: avoid creating orphan object in Client::check\_pool\_perm() ([issue#13782](#), [pr#6603](#), Yan, Zheng)
- client: avoid sending unnecessary FLUSHSNAP messages (Yan, Zheng)
- client: check if Fh is readable when processing a read ([issue#11517](#), [pr#7209](#), Yan, Zheng)
- client: close mds sessions in shutdown() ([pr#6269](#), John Spray)
- client: don't invalidate page cache when inode is no longer used ([pr#6380](#), Yan,

Zheng)

- client: don't mark\_down on command reply ([pr#6204](#), John Spray)
- client: drop prefix from ints ([pr#6275](#), John Coyle)
- client: exclude setfilelock when calculating oldest tid (Yan, Zheng)
- client: fix error handling in check\_pool\_perm (John Spray)
- client: flush kernel pagecache before creating snapshot ([issue#10436](#), [pr#7495](#), Yan, Zheng)
- client: fsync waits only for inode's caps to flush (Yan, Zheng)
- client: invalidate kernel dcache when cache size exceeds limits (Yan, Zheng)
- client: make fsync wait for unsafe dir operations (Yan, Zheng)
- client: modify a word in log ([pr#6906](#), YongQiang He)
- client: pin lookup dentry to avoid inode being freed (Yan, Zheng)
- client: properly trim unlinked inode ([issue#13903](#), [pr#7297](#), Yan, Zheng)
- client: removed unused Mutex from MetaRequest ([pr#7655](#), Greg Farnum)
- client: sys/file.h includes for flock operations ([pr#6282](#), John Coyle)
- client: use null snapc to check pool permission ([issue#13714](#), [pr#6497](#), Yan, Zheng)
- cls/cls\_rbd.cc: fix misused metadata\_name\_from\_key ([issue#13922](#), [pr#6661](#), Xiaoxi Chen)
- cls/cls\_rbd: pass string by reference ([pr#7232](#), Jeffrey Lu)
- cls\_hello: Fix grammatical error in description comment ([pr#7951](#), Brad Hubbard)
- cls\_journal: fix -EEXIST checking ([pr#8413](#), runsisi)
- cls\_rbd: add guards for error cases ([issue#14316](#), [issue#14317](#), [pr#7165](#), xie xingguo)
- cls\_rbd: change object\_map\_update to return 0 on success, add logging ([pr#6467](#), Douglas Fuller)
- cls\_rbd: enable object map checksums for object\_map\_save ([issue#14280](#), [pr#7149](#), Douglas Fuller)
- cls\_rbd: fix -EEXIST checking in cls::rbd::image\_set ([pr#8371](#), runsisi)
- cls\_rbd: fix the test for ceph-dencoder ([pr#7793](#), Kefu Chai)

- `cls_rbd`: `mirror_image_list` should return global image id ([pr#8297](#), Jason Dillaman)
- `cls_rbd`: mirroring directory ([issue#14419](#), [pr#7620](#), Josh Durgin)
- `cls_rbd`: pass `WILLNEED` fadvise flags during object map update ([issue#15332](#), [pr#8380](#), Jason Dillaman)
- `cls_rbd`: protect against excessively large object maps ([issue#15121](#), [pr#8099](#), Jason Dillaman)
- `cls_rbd`: `read_peers`: update `last_read` on next `cls_cxx_map_get_vals` ([pr#8374](#), Mykola Golub)
- `cls/rgw`: fix FTBFS ([pr#8142](#), Kefu Chai)
- `cls/rgw`: fix use of `timespan` ([issue#15181](#), [pr#8212](#), Yehuda Sadeh)
- `cmake`: add `common/fs_types.cc` to `libcommon` ([pr#7898](#), Orit Wasserman)
- `cmake`: Add `common/PluginRegistry.cc` to `CMakeLists.txt` ([pr#6805](#), Pete Zaitcev)
- `cmake`: Added new unittests to make check ([pr#7572](#), Ali Maredia)
- `cmake`: Add `ENABLE_GIT_VERSION` to avoid rebuilding ([pr#7171](#), Kefu Chai)
- `cmake`: add `ErasureCode.cc` to `jerasure` plugins ([pr#7808](#), Casey Bodley)
- `cmake`: add `FindOpenSSL.cmake` ([pr#8106](#), Marcus Watts, Matt Benjamin)
- `cmake`: add `KernelDevice.cc` to `libbos_srcs` ([pr#7507](#), Kefu Chai)
- `cmake`: add missing check for `HAVE_EXECINFO_H` ([pr#7270](#), Casey Bodley)
- `cmake`: add missing `librbd image_watcher` sources ([issue#14823](#), [pr#7717](#), Casey Bodley)
- `cmake`: add missing `librbd/MirrorWatcher.cc` and `librd/ObjectWatcher.cc` ([pr#8399](#), Orit Wasserman)
- `cmake`: add `nss` as a suffix for `pk11pub.h` ([pr#6556](#), Samuel Just)
- `cmake`: add `rgw_basic_types.cc` to `librgw.a` ([pr#6786](#), Orit Wasserman)
- `cmake`: add `StandardPolicy.cc` to `librbd` ([pr#8368](#), Kefu Chai)
- `cmake`: add `TracepointProvider.cc` to `libcommon` ([pr#6823](#), Orit Wasserman)
- `cmake`: avoid false-positive LDAP header detect ([pr#8100](#), Matt Benjamin)
- `cmake`: Build cython modules and change paths to `bin/`, `lib/` ([pr#8351](#), John Spray, Ali Maredia)

- cmake: check for libsnappy in default path also ([pr#7366](#), Kefu Chai)
- cmake: cleanups and more features from automake ([pr#7103](#), Casey Bodley, Ali Maredia)
- cmake: define STRERROR\_R\_CHAR\_P for GNU-specific strerror\_r ([pr#6751](#), Ilya Dryomov)
- cmake: detect bzip2 and lz4 ([pr#7126](#), Kefu Chai)
- cmake: feb5 ([pr#7541](#), Matt Benjamin)
- cmake: fix build with bluestore ([pr#7099](#), John Spray)
- cmake: fix files list ([pr#6539](#), Yehuda Sadeh)
- cmake: fix mrun to handle cmake build structure ([pr#8237](#), Orit Wasserman)
- cmake: fix paths to various EC source files ([pr#7748](#), Ali Maredia, Matt Benjamin)
- cmake: fix the build of test\_rados\_api\_list ([pr#8438](#), Kefu Chai)
- cmake: fix the build of tests ([pr#7523](#), Kefu Chai)
- cmake: fix the build on trusty ([pr#7249](#), Kefu Chai)
- cmake: For CMake version <= 2.8.11, use LINK\_PRIVATE and LINK\_PUBLIC ([pr#7474](#), Tao Chang)
- CMake: For CMake version <= 2.8.11, use LINK\_PRIVATE ([pr#8422](#), Haomai Wang)
- cmake: let ceph-client-debug link with tcmalloc ([pr#7314](#), Kefu Chai)
- cmake: librbd and libjournal build fixes ([pr#6557](#), Ilya Dryomov)
- cmake: made rocksdb an imported library ([pr#7131](#), Ali Maredia)
- cmake: no need to run configure from run-cmake-check.sh ([pr#6959](#), Orit Wasserman)
- cmake ([pr#7849](#), Ali Maredia)
- cmake: Remove duplicate find\_package libcurl line. ([pr#7972](#), Brad Hubbard)
- cmake: support ccache via a WITH\_CCACHE build option ([pr#6875](#), John Coyle)
- cmake: test\_build\_libcephfs needs \${ALLOC\_LIBS} ([pr#7300](#), Ali Maredia)
- cmake: update for recent librbd changes ([pr#6715](#), John Spray)
- cmake: update for recent rbd changes ([pr#6818](#), Mykola Golub)
- cmake: Use uname instead of arch. ([pr#6358](#), John Coyle)
- coc: fix typo in the apt-get command ([pr#6659](#), Chris Holcombe)

- common: add descriptions to perfcounters (Kiseleva Alyona)
- common: add generic plugin infrastructure ([pr#6696](#), Sage Weil)
- common: add latency perf counter for finisher ([pr#6175](#), Xinze Chi)
- common: add perf counter descriptions (Alyona Kiseleva)
- common/address\_help.cc: fix the leak in entity\_addr\_from\_url() ([issue#14132](#), [pr#6987](#), Qiankun Zheng)
- common: add thread names ([pr#5882](#), Igor Podoski)
- common: add zlib compression plugin ([pr#7437](#), Alyona Kiseleva, Kiseleva Alyona)
- common: admin socket commands for tcmalloc heap get/set operations ([pr#7512](#), Samuel Just)
- common: ake ceph\_time clocks work under BSD ([pr#7340](#), Adam C. Emerson)
- common: allow enable/disable of optracker at runtime ([pr#5168](#), Jianpeng Ma)
- common: Allow OPT\_INT settings with negative values ([issue#13829](#), [pr#7390](#), Brad Hubbard, Kefu Chai)
- common: assert: abort() rather than throw ([pr#6804](#), Adam C. Emerson)
- common: assert: \_\_STRING macro is not defined by musl libc. ([pr#6210](#), John Coyle)
- common/bit\_vector: use hard-coded value for block size ([issue#14747](#), [pr#7610](#), Jason Dillaman)
- common: buffer: add cached\_crc and cached\_crc\_adjust counts to perf dump ([pr#6535](#), Ning Yao)
- common: buffer/assert minor fixes ([pr#6990](#), Matt Benjamin)
- common: bufferlist performance tuning (Piotr Dalek, Sage Weil)
- common: buffer: put a guard for stat() syscall during read\_file ([pr#7956](#), xie xingguo)
- common: buffer: remove unneeded list destructor ([pr#6456](#), Michal Jarzabek)
- common/buffer: replace RWLock with spinlocks ([pr#7294](#), Piotr Dałek)
- common/ceph\_context.cc:fix order of initialisers ([pr#6838](#), Michal Jarzabek)
- common: change the type of counter total/unhealthy\_workers ([pr#7254](#), Guang Yang)
- common: default cluster name to config file prefix ([pr#7364](#), Javen Wu)
- common: Deprecate or free up a bunch of feature bits ([pr#8214](#), Samuel Just)

- common: detect overflow of int config values (#11484 Kefu Chai)
- common: Do not use non-portable constants in mutex\_debug ([pr#7766](#), Adam C. Emerson)
- common: don't reverse hobject\_t hash bits when zero ([pr#6653](#), Piotr Dałek)
- common: fix bit\_vector extent calc (#12611 Jason Dillaman)
- common: fix json parsing of utf8 (#7387 Tim Serong)
- common: fix leak of pthread\_mutexattr (#11762 Ketor Meng)
- common: fix LTTNG vs fork issue (Josh Durgin)
- common: fix OpTracker age histogram calculation ([pr#5065](#), Zhiqiang Wang)
- common: fix race during optracker switches between enabled/disabled mode ([pr#8330](#), xie xingguo)
- common: fix reset max in Throttle using perf reset command ([issue#13517](#), [pr#6300](#), Xinze Chi)
- common: fix throttle max change (Henry Chang)
- common: fix time\_t cast in decode ([issue#15330](#), [pr#8419](#), Adam C. Emerson)
- common/Formatter: avoid newline if there is no output ([pr#5351](#), Aran85)
- common: improve shared\_cache and simple\_cache efficiency with hash table ([pr#6909](#), Ning Yao)
- common/lockdep: increase max lock names ([pr#6961](#), Sage Weil)
- common: log: Assign LOG\_DEBUG priority to syslog calls ([issue#13993](#), [pr#6815](#), Brad Hubbard)
- common: log: predict log message buffer allocation size ([pr#6641](#), Adam Kupczyk)
- common: make mutex more efficient
- common: make work queue addition/removal thread safe (#12662 Jason Dillaman)
- common/MemoryModel: Added explicit feature check for mallinfo(). ([pr#6252](#), John Coyle)
- common: new timekeeping common code, and Objecter conversion ([pr#5782](#), Adam C. Emerson)
- common/obj\_bencher.cc: bump the precision of bandwidth field ([pr#8021](#), Piotr Dałek)
- common/obj\_bencher.cc: faster object name generation ([pr#7863](#), Piotr Dałek)

- common/obj\_bencher.cc: fix verification crashing when there's no objects ([pr#5853](#), Piotr Dałek)
- common/obj\_bencher.cc: make verify error fatal ([issue#14971](#), [pr#7897](#), Piotr Dałek)
- common: optimize debug logging code ([pr#6441](#), Adam Kupczyk)
- common: optimize debug logging ([pr#6307](#), Adam Kupczyk)
- common: optracker improvements (Zhiqiang Wang, Jianpeng Ma)
- common/page.cc: \_page\_mask has too many bits ([pr#7588](#), Dan Mick)
- common: perf counter for bufferlist history total alloc ([pr#6198](#), Xinze Chi)
- common: PriorityQueue tests (Kefu Chai)
- common: reduce CPU usage by making stringstream in stringify function thread local ([pr#6543](#), Evgeniy Firsov)
- common: re-enable backtrace support ([pr#6771](#), Jason Dillaman)
- common: set thread name from correct thread ([pr#7845](#), Igor Podoski)
- common: signal\_handler: added support for using reentrant strsignal() implementations vs. sys\_siglist[] ([pr#6796](#), John Coyle)
- common: snappy decompressor may assert when handling segmented input bufferlist ([issue#14400](#), [pr#7268](#), Igor Fedotov)
- common: some async compression infrastructure (Haomai Wang)
- common/str\_map: cleanup: replaced get\_str\_map() function overloading by using default parameters for delimiters ([pr#7266](#), Sahithi R V)
- common/strtol.cc: fix the coverity warnings ([pr#7967](#), Kefu Chai)
- common: SubProcess: Avoid buffer corruption when calling err() ([issue#15011](#), [pr#8054](#), Erwan Velu)
- common: SubProcess: fix multiple definition bug ([pr#6790](#), Yunchuan Wen)
- common: Thread: move copy constructor and assignment op ([pr#5133](#), Michal Jarzabek)
- common: time: have skewing-now call non-skewing now ([pr#7466](#), Adam C. Emerson)
- common/TrackedOp: fix inaccurate counting for slow requests ([issue#14804](#), [pr#7690](#), xie xingguo)
- common: unit test for interval\_set implementations ([pr#6](#), Igor Fedotov)

- common: use namespace instead of subclasses for buffer ([pr#6686](#), Michal Jarzabek)
- common: various fixes from SCA runs ([pr#7680](#), Danny Al-Gaaf)
- common: WorkQueue: new PointerWQ base class for ContextWQ ([issue#13636](#), [pr#6525](#), Jason Dillaman)
- compat: use prefixed typeof extension ([pr#6216](#), John Coyle)
- config: add \$data\_dir/config to config search path ([pr#7377](#), Sage Weil)
- config: complains when a setting is not tracked ([issue#11692](#), [pr#7085](#), Kefu Chai)
- config: fix osd\_crush\_initial\_weight ([pr#7975](#), You Ji)
- config: increase default async op threads ([pr#7802](#), Piotr Dałek)
- config\_opts: disable filestore throttle soft backoff by default ([pr#8265](#), Samuel Just)
- configure.ac: boost\_iostreams is required, not optional ([pr#7816](#), Hector Martin)
- configure.ac: macro fix ([pr#6769](#), Igor Podoski)
- configure.ac: make “-with-librocksdb-static” default to ‘check’ ([issue#14463](#), [pr#7317](#), Dan Mick)
- configure.ac: update help strings for cython ([pr#7856](#), Josh Durgin)
- configure: Add -D\_LARGEFILE64\_SOURCE to Linux build. ([pr#8402](#), Ira Cooper)
- configure: detect bz2 and lz4 ([issue#13850](#), [issue#13981](#), [pr#7030](#), Kefu Chai)
- correct radosgw-admin command ([pr#7006](#), YankunLi)
- crush: add -check to validate dangling names, max osd id (Kefu Chai)
- crush: add chooseleaf\_stable tunable ([pr#6572](#), Sangdi Xu, Sage Weil)
- crush: add safety assert ([issue#14496](#), [pr#7344](#), songbaisen)
- crush: cleanup, sync with kernel (Ilya Dryomov)
- crush: clean up whitespace removal ([issue#14302](#), [pr#7157](#), songbaisen)
- crush/CrushTester: check for overlapped rules ([pr#7139](#), Kefu Chai)
- crush/CrushTester: workaround a bug in boost::icl ([pr#7560](#), Kefu Chai)
- crush: fix cli tests for new crush tunables ([pr#8107](#), Sage Weil)
- crush: fix crash from invalid ‘take’ argument (#11602 Shiva Rkreddy, Sage Weil)
- crush: fix divide-by-2 in straw2 (#11357 Yann Dupont, Sage Weil)

- crush: fix error log ([pr#8430](#), Wei Jin)
- crush: fix has\_v4\_buckets (#11364 Sage Weil)
- crush: fix subtree base weight on adjust\_subtree\_weight (#11855 Sage Weil)
- crush: fix typo ([pr#8518](#), Wei Jin)
- crush: reply quickly from get\_immediate\_parent ([issue#14334](#), [pr#7181](#), song baisen)
- crush: respect default replicated ruleset config on map creation (Ilya Dryomov)
- crushtool: Don't crash when called on a file that isn't a crushmap ([issue#8286](#), [pr#8038](#), Brad Hubbard)
- crushtool: fix order of operations, usage (Sage Weil)
- crushtool: improve usage/tip messages ([pr#7142](#), xie xingguo)
- crushtool: set type 0 name "device" for -build option ([pr#6824](#), Sangdi Xu)
- crush: update tunable docs. change default profile to jewel ([pr#7964](#), Sage Weil)
- crush: validate bucket id before indexing buckets array ([issue#13477](#), [pr#6246](#), Sage Weil)
- crypto: fix NSS leak (Jason Dillaman)
- crypto: fix unbalanced init/shutdown (#12598 Zheng Yan)
- deb: fix rest-bench-dbg and ceph-test-dbg dependendies (Ken Dreyer)
- debian/changelog: Remove stray 'v' in version ([pr#7936](#), Dan Mick)
- debian/changelog: Remove stray 'v' in version ([pr#7938](#), Dan Mick)
- debian: include cpio in build-requiers ([pr#7533](#), Rémi BUISSON)
- debian: minor package reorg (Ken Dreyer)
- debian: package librgw\_file\* tests ([pr#7930](#), Ken Dreyer)
- debian: packaging fixes for jewel ([pr#7807](#), Ken Dreyer, Ali Maredia)
- debian/rpm split servers ([issue#10587](#), [pr#7746](#), Ken Dreyer)
- debian/rules: put init-ceph in /etc/init.d/ceph, not ceph-base ([issue#15329](#), [pr#8406](#), Dan Mick)
- deb, rpm: move ceph-objectstore-tool to ceph (Ken Dreyer)
- doc: add ceph-detect-init(8) source to dist tarball ([pr#7933](#), Ken Dreyer)

- doc: add cinder backend section to rbd-openstack.rst ([pr#7923](#), RustShen)
- doc: adding “-allow-shrink” in decreasing the size of the rbd block to distinguish from the increasing option ([pr#7020](#), Yehua)
- doc: add orphans commands to radosgw-admin(8) ([issue#14637](#), [pr#7518](#), Ken Dreyer)
- doc: add v0.80.11 to the release timeline ([pr#6658](#), Loic Dachary)
- doc: admin/build-doc: add lxml dependencies on debian ([pr#6610](#), Ken Dreyer)
- doc: admin/build-doc: make paths absolute ([pr#7119](#), Dan Mick)
- doc: amend Fixes instructions in SubmittingPatches ([pr#8312](#), Nathan Cutler)
- doc: amend the rados.8 ([pr#7251](#), Kefu Chai)
- doc/architecture.rst: remove redundant word “across” ([pr#8179](#), Zhao Junwang)
- doc/cephfs posix: update ([pr#6922](#), Sage Weil)
- doc: Clarify usage on starting single osd/mds/mon. ([pr#7641](#), Patrick Donnelly)
- doc: CodingStyle: fix broken URLs ([pr#6733](#), Kefu Chai)
- doc: correct typo ‘restared’ to ‘restarted’ ([pr#6734](#), Yilong Zhao)
- doc: detailed description of bugfixing workflow ([pr#7941](#), Nathan Cutler)
- doc/dev: add “Deploy a cluster for manual testing” section ([issue#15218](#), [pr#8228](#), Nathan Cutler)
- doc/dev: add section on interrupting a running suite ([pr#8116](#), Nathan Cutler)
- doc/dev: continue writing Testing in the cloud chapter ([pr#7960](#), Nathan Cutler)
- doc: dev: document ceph-qa-suite ([pr#6955](#), Loic Dachary)
- doc/dev/index: refactor/reorg ([pr#6792](#), Nathan Cutler)
- doc/dev/index.rst: begin writing Contributing to Ceph ([pr#6727](#), Nathan Cutler)
- doc/dev/index.rst: fix headings ([pr#6780](#), Nathan Cutler)
- doc/dev: integrate testing into the narrative ([pr#7946](#), Nathan Cutler)
- doc: dev: introduction to tests ([pr#6910](#), Loic Dachary)
- doc/dev: various refinements ([pr#7954](#), Nathan Cutler)
- doc: docuemnt object corpus generation (#11099 Alexis Normand)
- doc: document “readforward” and “readproxy” cache mode ([pr#7023](#), Kefu Chai)

- doc: document region hostnames (Robin H. Johnson)
- doc: download GPG key from download.ceph.com ([issue#13603](#), [pr#6384](#), Ken Dreyer)
- doc: draft notes for jewel ([pr#8211](#), Loic Dachary, Sage Weil)
- doc: file must be empty when writing layout fields of file use “setfattr” ([pr#6848](#), Cilang Zhao)
- doc: fix 0.94.4 and 0.94.5 ordering ([pr#7763](#), Loic Dachary)
- doc: Fixed incorrect name of a “List Multipart Upload Parts” Response Entity ([issue#14003](#), [pr#6829](#), Lenz Grimmer)
- doc: Fixes a CRUSH map step take argument ([pr#7327](#), Ivan Grcic)
- doc: Fixes a spelling error ([pr#6705](#), Jeremy Qian)
- doc: Fixes headline different font size and type ([pr#8328](#), scienceluo)
- doc: fix gender neutrality (Alexandre Maragone)
- doc: fixing image in section ERASURE CODING ([pr#7298](#), Rachana Patel)
- doc: fix install doc (#10957 Kefu Chai)
- doc: fix misleading configuration guide on cache tiering ([pr#7000](#), Yuan Zhou)
- doc: fix “mon osd down out subtree limit” option name ([pr#7164](#), François Lafont)
- doc: fix outdated content in cache tier ([pr#6272](#), Yuan Zhou)
- doc: fix S3 C# example ([pr#7027](#), Dunrong Huang)
- doc: fix sphinx issues (Kefu Chai)
- doc: fix typo, duplicated content etc. for Jewel release notes ([pr#8342](#), xie xingguo)
- doc: fix typo in cephfs/quota ([pr#6745](#), Drunkard Zhang)
- doc: fix typo, indentation etc. ([pr#7829](#), xie xingguo)
- doc: fix typo in developer guide ([pr#6943](#), Nathan Cutler)
- doc: fix typo ([pr#7004](#), tianqing)
- doc: fix wrong type of hyphen ([pr#8252](#), xie xingguo)
- doc: initial draft of RBD mirroring admin documentation ([issue#15041](#), [pr#8169](#), Jason Dillaman)
- doc: INSTALL redirect to online documentation ([pr#6749](#), Loic Dachary)

- doc: little improvements for troubleshooting scrub issues ([pr#6827](#), Mykola Golub)
- doc: man page updates (Kefu Chai)
- doc: mds data structure docs (Yan, Zheng)
- doc: misc updates (Fracois Lafont, Ken Dreyer, Kefu Chai, Owen Synge, Gael Fenet-Garde, Loic Dachary, Yannick Atchy-Dalama, Jiaying Ren, Kevin Caradant, Robert Maxime, Nicolas Yong, Germain Chipaux, Arthur Gorjux, Gabriel Sentucq, Clement Lebrun, Jean-Remi Deveaux, Clair Massot, Robin Tang, Thomas Laumondais, Jordan Dorne, Yuan Zhou, Valentin Thomas, Pierre Chaumont, Benjamin Troquereau, Benjamin Sesia, Vikhyat Umrao, Nilamdyuti Goswami, Vartika Rai, Florian Haas, Loic Dachary, Simon Guinot, Andy Allan, Alistair Israel, Ken Dreyer, Robin Rehu, Lee Revell, Florian Marsylle, Thomas Johnson, Bosse Klykken, Travis Rhoden, Ian Kelling)
- doc: Modified a note section in rbd-snapshot doc. ([pr#6908](#), Nilamdyuti Goswami)
- doc: note that cephfs auth stuff is new in jewel ([pr#6858](#), John Spray)
- doc: osd-config Add Configuration Options for op queue. ([pr#7837](#), Robert LeBlanc)
- doc: osd: s/schedued/scheduled/ ([pr#6872](#), Loic Dachary)
- doc/rados/api/librados-intro.rst: fix typo ([pr#7879](#), xie xingguo)
- doc/rados/operations/crush: fix the formatting ([pr#8306](#), Kefu Chai)
- doc: release-notes: draft v0.80.11 release notes ([pr#6374](#), Loic Dachary)
- doc: release-notes: draft v10.0.0 release notes ([pr#6666](#), Loic Dachary)
- doc/release-notes: fix indents ([pr#8345](#), Kefu Chai)
- doc/release-notes: v9.1.0 ([pr#6281](#), Loic Dachary)
- doc/releases-notes: fix build error ([pr#6483](#), Kefu Chai)
- doc: Remove Ceph Monitors do lots of fsync() ([issue#15288](#), [pr#8327](#), Vikhyat Umrao)
- doc: remove redundant space in ceph-authtool/monmaptool doc ([pr#7244](#), Jiaying Ren)
- doc: remove toctree items under Create CephFS ([pr#6241](#), Jevon Qiao)
- doc: remove unnecessary period in headline ([pr#6775](#), Marc Koderer)
- doc: rename the “Create a Ceph User” section and add verbage about... ([issue#13502](#), [pr#6297](#), ritz303)
- doc: revise SubmittingPatches ([pr#7292](#), Kefu Chai)

- doc: rgw admin uses “region list” not “regions list” ([pr#8517](#), Kris Jurka)
- doc: rgw explain keystone’s verify ssl switch ([pr#7862](#), Abhishek Lekshmanan)
- doc: rgw: port changes from downstream to upstream ([pr#7264](#), Bara Ancincova)
- doc: rgw\_region\_root\_pool option should be in [global] ([issue#15244](#), [pr#8271](#), Vikhyat Umrao)
- doc: rst style fix for pools document ([pr#6816](#), Drunkard Zhang)
- doc: script and guidelines for mirroring Ceph ([pr#7384](#), Wido den Hollander)
- docs: Fix styling of newly added mirror docs ([pr#6127](#), Wido den Hollander)
- doc: small fixes ([pr#7813](#), xiexingguo)
- doc: standardize @param (not @parma, @parmam, @params) ([pr#7714](#), Nathan Cutler)
- doc: SubmittingPatches: there is no next; only jewel ([pr#6811](#), Nathan Cutler)
- doc: swift tempurls (#10184 Abhishek Lekshmanan)
- doc: switch doxygen integration back to breathe (#6115 Kefu Chai)
- doc, tests: update all <http://ceph.com/> to download.ceph.com ([pr#6435](#), Alfredo Deza)
- doc: Update ceph-disk manual page with new feature deactivate/destroy. ([pr#6637](#), Vicente Cheng)
- doc: Updated CloudStack RBD documentation ([pr#8308](#), Wido den Hollander)
- doc: update doc for with new pool settings ([pr#5951](#), Guang Yang)
- doc: Updated the rados command man page to include the -run-name opt... ([issue#12899](#), [pr#5900](#), ritz303)
- doc: update infernalis release notes ([pr#6575](#), vasukulkarni)
- doc: Update list of admin/build-doc dependencies ([issue#14070](#), [pr#6934](#), Nathan Cutler)
- doc: update radosgw-admin example ([pr#6256](#), YankunLi)
- doc: update release schedule docs (Loic Dachary)
- doc: update the OS recommendations for newer Ceph releases ([pr#6355](#), ritz303)
- doc: use ‘ceph auth get-or-create’ for creating RGW keyring ([pr#6930](#), Wido den Hollander)
- doc: very basic doc on mstart ([pr#8207](#), Abhishek Lekshmanan)

- drop envz.h includes ([pr#6285](#), John Coyle)
- erasure-code: cleanup (Kefu Chai)
- erasure-code: improve tests (Loic Dachary)
- erasure-code: shec: fix recovery bugs (Takanori Nakao, Shotaro Kawaguchi)
- erasure-code: update ISA-L to 2.13 (Yuan Zhou)
- fix FTBFS introduced by d0af316 ([pr#7792](#), Kefu Chai)
- fix: use right init\_flags to finish CephContext ([pr#6549](#), Yunchuan Wen)
- fs: be more careful about the “mds setmap” command to prevent breakage ([issue#14380](#), [pr#7262](#), Yan, Zheng)
- ghobject\_t: use # instead of ! as a separator ([pr#8055](#), Sage Weil)
- global: do not start two daemons with a single pid-file ([issue#13422](#), [pr#7075](#), shun song)
- global: do not start two daemons with a single pid-file (part 2) ([issue#13422](#), [pr#7463](#), Loic Dachary)
- global/global\_init: expand metavariables in setuser\_match\_path ([issue#15365](#), [pr#8433](#), Sage Weil)
- global/signal\_handler: print thread name in signal handler ([pr#8177](#), Jianpeng Ma)
- gmock: switch to submodule (Danny Al-Gaaf, Loic Dachary)
- hadoop: add terasort test (Noah Watkins)
- helgrind: additional race conditionslibrbd: journal replay should honor inter-event dependencies ([pr#7274](#), Jason Dillaman)
- helgrind: fix real (and imaginary) race conditions ([issue#14163](#), [pr#7208](#), Jason Dillaman)
- include/encoding: do not try to be clever with list encoding ([pr#7913](#), Sage Weil)
- init-ceph: do umount when the path exists. ([pr#6866](#), Xiaoxi Chen)
- init-ceph.in: allow case-insensitive true in `osd crush update on start`` ([pr#7943](#), Eric Cook)
- init-ceph.in: skip ceph-disk if it is not present ([issue#10587](#), [pr#7286](#), Ken Dreyer)
- init-ceph: use getopt to make option processing more flexible ([issue#3015](#), [pr#6089](#), Nathan Cutler)

- init-radosgw: merge with sysv version; fix enumeration (Sage Weil)
- java: fix libcephfs bindings (Noah Watkins)
- journal: async methods to (un)register and update client ([pr#7832](#), Mykola Golub)
- journal: disconnect watch after watch error ([issue#14168](#), [pr#7113](#), Jason Dillaman)
- journal: fire replay complete event after reading last object ([issue#13924](#), [pr#6762](#), Jason Dillaman)
- journal: fix final result for JournalTrimmer::C\_RemoveSet ([pr#8516](#), runsisi)
- journal: fix race condition between Future and journal shutdown ([issue#15364](#), [pr#8477](#), Jason Dillaman)
- journal: flush commit position on metadata shutdown ([pr#7385](#), Mykola Golub)
- journal: improve commit position tracking ([pr#7776](#), Jason Dillaman)
- journal: incremental improvements and fixes ([pr#6552](#), Mykola Golub)
- journal: prevent race injecting new records into overflowed object ([issue#15202](#), [pr#8220](#), Jason Dillaman)
- journal: reset commit\_position\_task\_ctx pointer after task complete ([pr#7480](#), Mykola Golub)
- journal: re-use common threads between journalers ([pr#7906](#), Jason Dillaman)
- journal: support replaying beyond skipped splay objects ([pr#6687](#), Jason Dillaman)
- krbd: remove deprecated -quiet param from udevadm ([issue#13560](#), [pr#6394](#), Jason Dillaman)
- kv: fix bug in kv key optimization ([pr#6511](#), Sage Weil)
- kv: implement value\_as\_ptr() and use it in .get() ([pr#7052](#), Piotr Dałek)
- kv/KineticStore: fix broken split\_key ([pr#6574](#), Haomai Wang)
- kv: optimize and clean up internal key/value interface ([pr#6312](#), Piotr Dałek, Sage Weil)
- libcephfs: add pread, pwrite (Jevon Qiao)
- libcephfs,ceph-fuse: cache cleanup (Zheng Yan)
- libcephfs,ceph-fuse: fix request resend on cap reconnect (#10912 Yan, Zheng)
- libcephfs: fix python tests and fix.getcwd on missing dir ([pr#7901](#), John Spray)

- libcephfs: Improve portability by replacing loff\_t type usage with off\_t ([pr#6301](#), John Coyle)
- libcephfs: only check file offset on glibc platforms ([pr#6288](#), John Coyle)
- libcephfs: update LIBCEPHFS\_VERSION to indicate the interface was changed ([pr#7551](#), Jevon Qiao)
- librados: add config observer (Alistair Strachan)
- librados: add c++ style osd/pg command interface ([pr#6893](#), Yunchuan Wen)
- librados: add FULL\_TRY and FULL\_FORCE flags for dealing with full clusters or pools (Sage Weil)
- librados: add src\_fadvise\_flags for copy-from (Jianpeng Ma)
- librados: aix gcc librados port ([pr#6675](#), Rohan Mars)
- librados: avoid malloc(0) (which can return NULL on some platforms) ([issue#13944](#), [pr#6779](#), Dan Mick)
- librados: cancel aio notification linger op upon completion ([pr#8102](#), Jason Dillaman)
- librados: check connection state in rados\_monitor\_log ([issue#14499](#), [pr#7350](#), David Disseldorp)
- librados: clean up Objecter.h ([pr#6731](#), Jie Wang)
- librados: define C++ flags from C constants (Josh Durgin)
- librados: detect laggy ops with objecter\_timeout, not osd\_timeout ([pr#7629](#), Greg Farnum)
- librados: do cleanup ([pr#6488](#), xie xingguo)
- librados: do not clear handle for aio\_watch() ([pr#7771](#), xie xingguo)
- librados: fadvise flags per op (Jianpeng Ma)
- librados: fix examples/librados/Makefile error. ([pr#6320](#), You Ji)
- librados: fix last\_force\_resent handling (#11026 Jianpeng Ma)
- librados: fix memory leak from C\_TwoContexts (Xiong Yiliang)
- librados: fix notify completion race (#13114 Sage Weil)
- librados: fix pool alignment API overflow issue ([issue#13715](#), [pr#6489](#), xie xingguo)
- librados: fix potential null pointer access when do pool\_snap\_list ([issue#13639](#),

pr#6422, xie xingguo)

- librados: fix PromoteOn2ndRead test for EC (pr#6373, Sage Weil)
- librados: fix rare race where pool op callback may hang forever (issue#13642, pr#6426, xie xingguo)
- librados: fix several flaws introduced by the enumeration\_objects API (issue#14299, issue#14301, issue#14300, pr#7156, xie xingguo)
- librados: fix striper when stripe\_count = 1 and stripe\_unit != object\_size (#11120 Yan, Zheng)
- librados: fix test failure with new aio watch/unwatch API (pr#7824, Jason Dillaman)
- librados: implement async watch/unwatch (pr#7649, Haomai Wang)
- librados: include/rados/librados.h: fix typo (pr#6741, Nathan Cutler)
- librados: init crush\_location from config file. (issue#13473, pr#6243, Wei Luo)
- librados, libcephfs: randomize client nonces (Josh Durgin)
- librados: mix lock cycle (un)registering asok commands (pr#7581, John Spray)
- librados: move to c++11 concurrency types (pr#5931, Adam C. Emerson)
- librados: new style (sharded) object listing (pr#6405, John Spray, Sage Weil)
- librados: op perf counters (John Spray)
- librados: potential null pointer access in list\_(n)objects (issue#13822, pr#6639, xie xingguo)
- librados: pybind: fix binary omap values (Robin H. Johnson)
- librados: pybind: fix write() method return code (Javier Guerra)
- librados: race condition on aio\_notify completion handling (pr#7864, Jason Dillaman)
- librados: remove duplicate definitions for rados pool\_stat\_t and cluster\_stat\_t (pr#7330, Igor Fedotov)
- librados: respect default\_crush\_ruleset on pool\_create (#11640 Yuan Zhou)
- librados: Revert “rados: Add new field flags for ceph\_osd\_op.copy\_get.” (pr#8486, Sage Weil)
- librados: shutdown finisher in a more graceful way (pr#7519, xie xingguo)
- librados: Solaris port (pr#6416, Rohan Mars)

- librados: stat2 with higher time precision ([pr#7915](#), Yehuda Sadeh, Matt Benjamin)
- librados: Striper: Fix incorrect push\_front -> append\_zero change ([pr#7578](#), Haomai Wang)
- libradosstriper: fix leak (Danny Al-Gaaf)
- librados\_test\_stub: protect against notify/unwatch race ([pr#7540](#), Jason Dillaman)
- librados: wrongly passed in argument for stat command ([issue#13703](#), [pr#6476](#), xie xingguo)
- librbd: add const for single-client-only features (Josh Durgin)
- librbd: add deep-flatten operation (Jason Dillaman)
- librbd: add purge\_on\_error cache behavior (Jianpeng Ma)
- librbd: allocate new journal tag after acquiring exclusive lock ([pr#7884](#), Jason Dillaman)
- librbd: allow additional metadata to be stored with the image (Haomai Wang)
- librbd: API: async open and close ([issue#14264](#), [pr#7259](#), Mykola Golub)
- librbd: automatically flush IO after blocking write operations ([issue#13913](#), [pr#6742](#), Jason Dillaman)
- librbd: avoid blocking aio API methods (#11056 Jason Dillaman)
- librbd: Avoid create two threads per image ([pr#7400](#), Haomai Wang)
- librbd: avoid throwing error if mirroring is unsupported ([pr#8417](#), Jason Dillaman)
- librbd: better handling for dup flatten requests (#11370 Jason Dillaman)
- librbd: better handling of exclusive lock transition period ([pr#7204](#), Jason Dillaman)
- librbd: block maintenance ops until after journal is ready ([issue#14510](#), [pr#7382](#), Jason Dillaman)
- librbd: block read requests until journal replayed ([pr#7627](#), Jason Dillaman)
- librbd: cancel in-flight ops on watch error (#11363 Jason Dillaman)
- librbd: check for presence of journal before attempting to remove ([issue#13912](#), [pr#6737](#), Jason Dillaman)
- librbd: clear error when older OSD doesn't support image flags ([issue#14122](#), [pr#7035](#), Jason Dillaman)

- librbd: correct include guard in RenameRequest.h ([pr#7143](#), Jason Dillaman)
- librbd: correct issues discovered during teuthology testing ([issue#14108](#), [issue#14107](#), [pr#6974](#), Jason Dillaman)
- librbd: correct issues discovered via valgrind memcheck ([pr#8132](#), Jason Dillaman)
- librbd: correct issues discovered when cache is disabled ([issue#14123](#), [pr#6979](#), Jason Dillaman)
- librbd: correct race conditions discovered during unit testing ([issue#14060](#), [pr#6923](#), Jason Dillaman)
- librbd: deadlock while attempting to flush AIO requests ([issue#13726](#), [pr#6508](#), Jason Dillaman)
- librbd: default new images to format 2 (#11348 Jason Dillaman)
- librbd: differentiate journal replay flush vs shut down ([pr#7698](#), Jason Dillaman)
- librbd: disable copy-on-read when not exclusive lock owner ([issue#14167](#), [pr#7129](#), Jason Dillaman)
- librbd: disable image mirroring when image is removed ([issue#15265](#), [pr#8375](#), Ricardo Dias)
- librbd: disallow unsafe rbd\_op\_threads values ([issue#15034](#), [pr#8459](#), Josh Durgin)
- librbd: do not ignore self-managed snapshot release result ([issue#14170](#), [pr#7043](#), Jason Dillaman)
- librbd: enable/disable image mirroring automatically for pool mode ([issue#15143](#), [pr#8204](#), Ricardo Dias)
- librbd: ensure copy-on-read requests are complete prior to closing parent image ([pr#6740](#), Jason Dillaman)
- librbd: ensure librados callbacks are flushed prior to destroying ([issue#14092](#), [pr#7040](#), Jason Dillaman)
- librbd: exit if parent's snap is gone during clone ([issue#14118](#), [pr#6968](#), xie xingguo)
- librbd: fadvise for copy, export, import (Jianpeng Ma)
- librbd: fast diff implementation that leverages object map (Jason Dillaman)
- librbd: fix enable objectmap feature issue ([issue#13558](#), [pr#6339](#), xinxin shu)
- librbd: fix fast diff bugs (#11553 Jason Dillaman)
- librbd: fix image format detection (Zhiqiang Wang)

- librbd: fix internal handling of dynamic feature updates ([pr#7299](#), Jason Dillaman)
- librbd: fix journal iohint ([pr#6917](#), Jianpeng Ma)
- librbd: fix known test case race condition failures ([issue#13969](#), [pr#6800](#), Jason Dillaman)
- librbd: fix lock ordering issue (#11577 Jason Dillaman)
- librbd: fix merge-diff for >2GB diff-files ([issue#14030](#), [pr#6889](#), Yunchuan Wen)
- librbd: fix potential memory leak ([issue#14332](#), [issue#14333](#), [pr#7174](#), xie xingguo)
- librbd: fix reads larger than the cache size (Lu Shi)
- librbd: fix snap\_exists API return code overflow ([issue#14129](#), [pr#6986](#), xie xingguo)
- librbd: fix snapshot creation when other snap is active (#11475 Jason Dillaman)
- librbd: fix state machine race conditions during shut down ([pr#7761](#), Jason Dillaman)
- librbd: fix test case race condition for journaling ops ([pr#6877](#), Jason Dillaman)
- librbd: fix tracepoint parameter in diff\_iterate ([pr#6892](#), Yunchuan Wen)
- librbd: flatten/copyup fixes (Jason Dillaman)
- librbd: flush and invalidate cache via admin socket ([issue#2468](#), [pr#6453](#), Mykola Golub)
- librbd: handle NOCACHE fadvise flag (Jinapeng Ma)
- librbd: handle unregistering the image watcher when disconnected ([pr#8094](#), Jason Dillaman)
- librbd: image refresh code paths converted to async state machines ([pr#6859](#), Jason Dillaman)
- librbd: include missing header for bool type ([pr#6798](#), Mykola Golub)
- librbd: initial collection of state machine unit tests ([pr#6703](#), Jason Dillaman)
- librbd: integrate journaling for maintenance operations ([pr#6625](#), Jason Dillaman)
- librbd: integrate journaling support for IO operations ([pr#6541](#), Jason Dillaman)
- librbd: integrate journal replay with fsx testing ([pr#7583](#), Jason Dillaman)
- librbd: journal framework for tracking exclusive lock transitions ([issue#13298](#),

pr#7529, Jason Dillaman)

- librbd: journaling-related lock dependency cleanup (pr#6777, Jason Dillaman)
- librbd: journal replay needs to support re-executing maintenance ops (issue#14822, pr#7785, Jason Dillaman)
- librbd: journal replay should honor inter-event dependencies (pr#7019, Jason Dillaman)
- librbd: journal shut down flush race condition (issue#14434, pr#7302, Jason Dillaman)
- librbd: lockdep, helgrind validation (Jason Dillaman, Josh Durgin)
- librbd: metadata filter fixes (Haomai Wang)
- librbd: misc aio fixes (#5488 Jason Dillaman)
- librbd: misc rbd fixes (#11478 #11113 #11342 #11380 Jason Dillaman, Zhiqiang Wang)
- librbd: new diff\_iterate2 API (Jason Dillaman)
- librbd: not necessary to hold owner\_lock while releasing snap id (issue#13914, pr#6736, Jason Dillaman)
- librbd: object map rebuild support (Jason Dillaman)
- librbd: only send signal when AIO completions queue empty (pr#6729, Jianpeng Ma)
- librbd: only update image flags while hold exclusive lock (#11791 Jason Dillaman)
- librbd: optionally disable allocation hint (Haomai Wang)
- librbd: optionally validate new RBD pools for snapshot support (issue#13633, pr#6925, Jason Dillaman)
- librbd: partial revert of commit 9b0e359 (issue#13969, pr#6789, Jason Dillaman)
- librbd: perf counters might not be initialized on error (issue#13740, pr#6523, Jason Dillaman)
- librbd: perf section name: use hyphen to separate components (issue#13719, pr#6516, Mykola Golub)
- librbd: prevent race between resize requests (#12664 Jason Dillaman)
- librbd: properly handle replay of snap remove RPC message (issue#14164, pr#7042, Jason Dillaman)
- librbd: readahead fixes (Zhiqiang Wang)

- librbd: reduce mem copies to user-buffer during read ([pr#7548](#), Jianpeng Ma)
- librbd: reduce verbosity of common error condition logging ([issue#14234](#), [pr#7114](#), Jason Dillaman)
- librbd: refresh image if required before replaying journal ops ([issue#14908](#), [pr#7978](#), Jason Dillaman)
- librbd: remove canceled tasks from timer thread ([issue#14476](#), [pr#7329](#), Douglas Fuller)
- librbd: remove duplicate read\_only test in librbd::async\_flatten ([pr#5856](#), runsisi)
- librbd: remove last synchronous librados calls from open/close state machine ([pr#7839](#), Jason Dillaman)
- librbd: replaying a journal op post-refresh requires locking ([pr#8028](#), Jason Dillaman)
- librbd: resize should only update image size within header ([issue#13674](#), [pr#6447](#), Jason Dillaman)
- librbd: retrieve image name when opening by id ([pr#7736](#), Mykola Golub)
- librbd: return error if we fail to delete object\_map head object ([issue#14098](#), [pr#6958](#), xie xingguo)
- librbd: return result code from close (#12069 Jason Dillaman)
- librbd: Revert “librbd: use task finisher per CephContext” ([issue#14780](#), [pr#7667](#), Josh Durgin)
- librbd: send notifications for mirroring status updates ([pr#8355](#), Jason Dillaman)
- librbd: several race conditions discovered under single CPU environment ([pr#7653](#), Jason Dillaman)
- librbd: simplify IO method signatures for 32bit environments ([pr#6700](#), Jason Dillaman)
- librbd: small fixes for error messages and readahead counter ([issue#14127](#), [pr#6983](#), xie xingguo)
- librbd: start perf counters after id is initialized ([issue#13720](#), [pr#6494](#), Mykola Golub)
- librbd: store metadata, including config options, in image (Haomai Wang)
- librbd: support eventfd for AIO completion notifications ([pr#5465](#), Haomai Wang)
- librbd: tolerate old osds when getting image metadata (#11549 Jason Dillaman)

- librbd: truncate does not need to mark the object as existing in the object map ([issue#14789](#), [pr#7772](#), xinxin shu)
- librbd: uninitialized state in snap remove state machine ([pr#6982](#), Jason Dillaman)
- librbd: update of mirror pool mode and mirror peer handling ([pr#7718](#), Jason Dillaman)
- librbd: use async librados notifications ([pr#7668](#), Jason Dillaman)
- librbd: use write\_full when possible (Zhiqiang Wang)
- log: do not repeat errors to stderr ([issue#14616](#), [pr#7983](#), Sage Weil)
- log: fix data corruption race resulting from log rotation (#12465 Samuel Just)
- log: fix stack overflow when flushing large log lines ([issue#14707](#), [pr#7599](#), Igor Fedotov)
- logrotate.d: prefer service over invoke-rc.d (#11330 Win Hierman, Sage Weil)
- log: segv in a portable way ([issue#14856](#), [pr#7790](#), Kefu Chai)
- log: use delete[] ([pr#7904](#), Sage Weil)
- mailmap: add UMCLOUD affiliation ([pr#6820](#), Jiaying Ren)
- mailmap for 10.0.4 ([pr#7932](#), Abhishek Lekshmanan)
- mailmap: hange organization for Dongmao Zhang ([pr#7173](#), Dongmao Zhang)
- mailmap: Igor Podoski affiliation ([pr#7219](#), Igor Podoski)
- mailmap: Jewel updates ([pr#6750](#), Abhishek Lekshmanan)
- mailmap: modify member info ([pr#6468](#), Xiaowei Chen)
- mailmap: revise organization ([pr#6519](#), Li Wang)
- mailmap: Ubuntu Kylin name changed to Kylin Cloud ([pr#6532](#), Loic Dachary)
- mailmap: update .organizationmap ([pr#6565](#), chenji-kael)
- mailmap update ([pr#7210](#), M Ranga Swami Reddy)
- mailmap update ([pr#8522](#), M Ranga Swami Reddy)
- mailmap: updates for infernalis. ([pr#6495](#), Yann Dupont)
- mailmap: updates ([pr#6258](#), M Ranga Swami Reddy)
- mailmap: updates ([pr#6594](#), chenji-kael)

- mailmap updates ([pr#6992](#), Loic Dachary)
- mailmap updates ([pr#7189](#), Loic Dachary)
- mailmap updates ([pr#7528](#), Yann Dupont)
- mailmap updates ([pr#8256](#), Loic Dachary)
- mailmap: Xie Xingguo affiliation ([pr#6409](#), Loic Dachary)
- Makefile-env.am: set a default for CEPH\_BUILD\_VIRTUALENV (part 2) ([pr#8320](#), Loic Dachary)
- makefile: fix rbdmap manpage ([pr#8310](#), Kefu Chai)
- makefile: remove libedit from libclient.la ([pr#7284](#), Kefu Chai)
- makefiles: remove bz2-dev from dependencies ([issue#13981](#), [pr#6939](#), Piotr Dałek)
- man/8/ceph-disk: fix formatting issue ([pr#8003](#), Sage Weil)
- man/8/ceph-disk: fix formatting issue ([pr#8012](#), Sage Weil)
- man: document listwatchers cmd in “rados” manpage ([pr#7021](#), Kefu Chai)
- mdsa: A few more snapshot fixes, mostly around snapshotted inode/dentry tracking ([pr#7798](#), Yan, Zheng)
- mds: Add cmapv to ESessions default constructor initializer list ([pr#8403](#), John Coyle)
- mds: add ‘damaged’ state to MDSMap (John Spray)
- mds: add nicknames for perfcounters (John Spray)
- mds: add ‘p’ flag in auth caps to control setting pool in layout ([pr#6567](#), John Spray)
- mds: advance clientreplay when replying ([issue#14357](#), [pr#7216](#), John Spray)
- mds: allow client to request caps when opening file ([issue#14360](#), [pr#7952](#), Yan, Zheng)
- mds: avoid emitting cap warnigns before evicting session (John Spray)
- mds: avoid getting stuck in XLOCKDONE (#11254 Yan, Zheng)
- mds, client: add namespace to file\_layout\_t (previously ceph\_file\_layout) ([pr#7098](#), Yan, Zheng, Sage Weil)
- mds, client: fix locking around handle\_conf\_change ([issue#14365](#), [issue#14374](#), [pr#7312](#), John Spray)

- mds: disable problematic rstat propagation into snap parents (Yan, Zheng)
- mds: do not add snapped items to bloom filter (Yan, Zheng)
- mds: don't double-shutdown the timer when suiciding ([issue#14697](#), [pr#7616](#), Greg Farnum)
- mds: expose frags via asok (John Spray)
- mds: expose state of recovery to status ASOK command ([issue#14146](#), [pr#7068](#), Yan, Zheng)
- mds: Extend the existing pool access checking to include specific RADOS namespaces. ([pr#8444](#), Yan, Zheng)
- mds: filelock deadlock ([pr#7713](#), Yan, Zheng)
- mds: fix client capabilities during reconnect (client.XXXX isn't responding to mclientcaps(revoke)) ([issue#11482](#), [pr#6432](#), Yan, Zheng)
- mds: fix client cap/message replay order on restart ([issue#14254](#), [issue#13546](#), [pr#7199](#), Yan, Zheng)
- mds: fix expected holes in journal objects (#13167 Yan, Zheng)
- mds: fix file\_layout\_t legacy encoding snafu ([pr#8455](#), Sage Weil)
- mds: fix fsmap decode ([pr#8063](#), Greg Farnum)
- mds: fix FSMap upgrade with daemons in the map ([pr#8073](#), John Spray, Greg Farnum)
- mds: fix handling for missing mydir dirfrag (#11641 John Spray)
- mds: fix inode\_t::compare() ([issue#15038](#), [pr#8014](#), Yan, Zheng)
- mds: fix integer truncation on large client ids (Henry Chang)
- mds: fix mydir replica issue with shutdown (#10743 John Spray)
- mds: fix out-of-order messages (#11258 Yan, Zheng)
- mds: fix rejoin (Yan, Zheng)
- mds: fix scrub\_path ([pr#6684](#), John Spray)
- mds: fix setting entire file layout in one setxattr (John Spray)
- mds: fix setvxattr (broken in a536d114) ([issue#14029](#), [pr#6941](#), John Spray)
- mds: fix shutdown (John Spray)
- mds: fix shutdown with strays (#10744 John Spray)

- mds: fix SnapServer crash on deleted pool (John Spray)
- mds: fix snapshot bugs (Yan, Zheng)
- mds: fix standby replay thread creation ([issue#14144](#), [pr#7132](#), John Spray)
- mds: fix stray handling (John Spray)
- mds: fix stray purging in 'stripe\_count > 1' case ([issue#15050](#), [pr#8040](#), Yan, Zheng)
- mds: fix stray reintegration (Yan, Zheng)
- mds: fix suicide beacon (John Spray)
- mds: flush immediately in do\_open\_truncate (#11011 John Spray)
- mds: function parameter 'df' should be passed by reference ([pr#7490](#), Na Xie)
- mds: handle misc corruption issues (John Spray)
- mds: implement snapshot rename ([pr#5645](#), xinxin shu)
- mds: improve dump methods (John Spray)
- mds: judgment added to avoid the risk of visiting the NULL pointer ([pr#7358](#), Kongming Wu)
- mds: many fixes (Yan, Zheng, John Spray, Greg Farnum)
- mds: many snapshot and stray fixes (Yan, Zheng)
- mds: messages/MOSDOp: cast in assert to eliminate warnings ([issue#13625](#), [pr#6414](#), David Zafman)
- mds: misc fixes (Jianpeng Ma, Dan van der Ster, Zhang Zhi)
- mds: misc journal cleanups and fixes (#10368 John Spray)
- mds: misc repair improvements (John Spray)
- mds: misc snap fixes (Zheng Yan)
- mds: misc snapshot fixes (Yan, Zheng)
- mds: Multi-filesystem support ([issue#14952](#), [pr#6953](#), John Spray, Sage Weil)
- mds: new filtered MDS tell commands for sessions ([pr#6180](#), John Spray)
- mds: new SessionMap storage using omap (#10649 John Spray)
- mds: persist completed\_requests reliably (#11048 John Spray)
- mds: properly set STATE\_STRAY/STATE\_ORPHAN for stray dentry/inode ([issue#13777](#),

[pr#6553](#), Yan, Zheng)

- mds: Protect a number of unstable/experimental features behind durable flags ([pr#8383](#), Greg Farnum)
- mds: reduce memory consumption (Yan, Zheng)
- mds: repair the command option “-hot-standby” ([pr#6454](#), Wei Feng)
- mds: respawn instead of suicide on blacklist (John Spray)
- mds: ScrubStack and “tag path” command ([pr#5662](#), Yan, Zheng, John Spray, Greg Farnum)
- mds: separate safe\_pos in Journaler (#10368 John Spray)
- mds/Session: use projected parent for auth path check ([issue#13364](#), [pr#6200](#), Sage Weil)
- mds: snapshot rename support (#3645 Yan, Zheng)
- mds: store layout on header object (#4161 John Spray)
- mds: tear down connections from tell commands ([issue#14048](#), [pr#6933](#), John Spray)
- mds: throttle purge stray operations (#10390 John Spray)
- mds: tolerate clock jumping backwards (#11053 Yan, Zheng)
- mds: warn when clients fail to advance oldest\_client\_tid (#10657 Yan, Zheng)
- mds: we should wait messenger when MDSDaemon suicide ([pr#6996](#), Wei Feng)
- messages/MOSDOp: clear reqid inc for v6 encoding ([issue#15230](#), [pr#8299](#), Sage Weil)
- Minor fixes around data scan in some scenarios ([pr#8115](#), Yan, Zheng)
- mirrors: Change contact e-mail address for se.ceph.com ([pr#8007](#), Wido den Hollander)
- mirrors: Updated scripts and documentation for mirrors ([pr#7847](#), Wido den Hollander)
- misc cleanups and fixes (Danny Al-Gaaf)
- misc coverity fixes (Danny Al-Gaaf)
- misc performance and cleanup (Nathan Cutler, Xinxin Shu)
- misc: use make\_shared while creating shared\_ptr ([pr#7769](#), Somnath Roy)
- mon: add an independent option for max election time ([pr#7245](#), Sangdi Xu)

- mon: add cache over MonitorDBStore (Kefu Chai)
- mon: add ‘mon\_metadata <id>’ command (Kefu Chai)
- mon: add ‘node ls ...’ command (Kefu Chai)
- mon: add NOFORWARD, OBSOLETE, DEPRECATE flags for mon commands (Joao Eduardo Luis)
- mon: add osd blacklist clear ([pr#6945](#), John Spray)
- mon: add PG count to ‘ceph osd df’ output (Michal Jarzabek)
- mon: add RAW USED column to ceph df detail ([pr#7087](#), Ruifeng Yang)
- mon: block ‘ceph osd pg-temp ...’ if pg\_temp update is already pending ([pr#6704](#), Sage Weil)
- mon: ‘ceph osd metadata’ can dump all osds (Haomai Wang)
- mon: clean up, reorg some mon commands (Joao Eduardo Luis)
- mon: cleanup set-quota error msg ([pr#7371](#), Abhishek Lekshmanan)
- monclient: avoid key renew storm on clock skew ([issue#12065](#), [pr#8258](#), Alexey Sheplyakov)
- monclient: flush\_log (John Spray)
- mon: compact full epochs also ([issue#14537](#), [pr#7396](#), Kefu Chai)
- mon: consider pool size when creating pool ([issue#14509](#), [pr#7359](#), songbaisen)
- mon: consider the pool size when setting pool crush rule ([issue#14495](#), [pr#7341](#), song baisen)
- mon: degrade a log message to level 2 ([pr#6929](#), Kongming Wu)
- mon: detect kv backend failures (Sage Weil)
- mon: disallow >2 tiers (#11840 Kefu Chai)
- mon: disallow ec pools as tiers (#11650 Samuel Just)
- mon: do not deactivate last mds (#10862 John Spray)
- mon: do not send useless pg\_create messages for split pgs ([pr#8247](#), Sage Weil)
- mon: don’t require OSD W for MRemoveSnaps ([issue#13777](#), [pr#6601](#), John Spray)
- mon: drop useless rank init assignment ([issue#14508](#), [pr#7321](#), huanwen ren)
- mon: enable ‘mon osd prime pg temp’ by default ([pr#7838](#), Robert LeBlanc)

- mon: fix average utilization calc for 'osd df' (Mykola Golub)
- mon: fix calculation of %USED (pr#7881, Adam Kupczyk)
- mon: fix ceph df pool available calculation for 0-weighted OSDs (pr#6660, Chengyuan Li)
- mon: fix coding-style on PG related Monitor files (pr#6881, Wido den Hollander)
- mon: fix CRUSH map test for new pools (Sage Weil)
- mon: fixes related to mondbstore->get() changes (pr#6564, Piotr Dałek)
- mon: fix keyring permissions (issue#14950, pr#7880, Owen Synge)
- mon: fix locking in preinit error paths (issue#14473, pr#7353, huanwen ren)
- mon: fix log dump crash when debugging (Mykola Golub)
- mon: fix mds beacon replies (#11590 Kefu Chai)
- mon: fix metadata update race (Mykola Golub)
- mon: fix min\_last\_epoch\_clean tracking (Kefu Chai)
- mon: fix monmap creation stamp (pr#7459, duanweijun)
- mon: fix 'pg ls' sort order, state names (#11569 Kefu Chai)
- mon: fix refresh (#11470 Joao Eduardo Luis)
- mon: fix reuse of osd ids (clear osd info on osd deletion) (issue#13988, pr#6900, Loic Dachary, Sage Weil)
- mon: fix routed\_request\_tids leak (pr#6102, Ning Yao)
- mon: fix sync of config-key data (pr#7363, Xiaowei Chen)
- mon: fix the can't change subscribe level bug in monitoring log (pr#7031, Zhiqiang Wang)
- mon: fix variance calc in 'osd df' (Sage Weil)
- mon: go into ERR state if multiple PGs are stuck inactive (issue#13923, pr#7253, Wido den Hollander)
- mon: improve callout to crushtool (Mykola Golub)
- mon: initialize last\_\* timestamps on new pgs to creation time (issue#14952, pr#7980, Sage Weil)
- mon: initialize recorded election epoch properly even when standalone (issue#13627, pr#6407, huanwen ren)

- mon: make blocked op messages more readable (Jianpeng Ma)
- mon: make clock skew checks sane ([issue#14175](#), [pr#7141](#), Joao Eduardo Luis)
- mon: make osd get pool 'all' only return applicable fields (#10891 Michal Jarzabek)
- mon: mark\_down\_pgs in lockstep with pg\_map's osdmap epoch ([pr#8208](#), Sage Weil)
- mon/MDSMonitor: add confirmation to "ceph mds rmfailed" ([issue#14379](#), [pr#7248](#), Yan, Zheng)
- mon/MDSMonitor.cc: properly note beacon when health metrics changes ([issue#14684](#), [pr#7757](#), Yan, Zheng)
- mon: misc scaling fixes (Sage Weil)
- mon: modify a dout level in OSDMonitor.cc ([pr#6928](#), Yongqiang He)
- mon/MonClient: avoid null pointer error when configured incorrectly ([issue#14405](#), [pr#7276](#), Bo Cai)
- mon/MonClient: fix shutdown race ([issue#13992](#), [pr#8335](#), Sage Weil)
- mon/monitor: some clean up ([pr#7520](#), huanwen ren)
- mon: MonmapMonitor: don't expose uncommitted state to client ([pr#6854](#), Joao Eduardo Luis)
- mon: normalize erasure-code profile for storage and comparison (Loic Dachary)
- mon: only send mon metadata to supporting peers (Sage Weil)
- mon: optionally specify osd id on 'osd create' (Mykola Golub)
- mon/OSDMonitor: osdmap laggy set a maximum limit for interval ([pr#7109](#), Zengran Zhang)
- mon: osd [test-]reweight-by-{pg,utilization} command updates ([pr#7890](#), Dan van der Ster, Sage Weil)
- mon: 'osd tree' fixes (Kefu Chai)
- mon: paxos is\_recovering calc error ([pr#7227](#), Weijun Duan)
- mon: periodic background scrub (Joao Eduardo Luis)
- mon/PGMap: show rd/wr iops separately in status reports ([pr#7072](#), Cilang Zhao)
- mon: PGMonitor: acting primary diff with cur\_stat, should not set pg to stale ([pr#7083](#), Xiaowei Chen)
- mon/PGMonitor: reliably mark PGs state ([pr#8089](#), Sage Weil)

- mon: PG Monitor should report waiting for backfill ([issue#12744](#), [pr#7398](#), Abhishek Lekshmanan)
- mon/pgmonitor: use appropriate forced conversions in get\_rule\_avail ([pr#7705](#), huanwen ren)
- mon: prevent bucket deletion when referenced by a crush rule (#11602 Sage Weil)
- mon: prevent pgp\_num > pg\_num (#12025 Xinxin Shu)
- mon: prevent pool with snapshot state from being used as a tier (#11493 Sage Weil)
- mon: prime pg\_temp when CRUSH map changes (Sage Weil)
- mon: reduce CPU and memory manager pressure of pg health check ([pr#7482](#), Piotr Dałek)
- mon: refine check\_remove\_tier checks (#11504 John Spray)
- mon: reject large max\_mds values (#12222 John Spray)
- mon: remove 'mds setmap' ([issue#15136](#), [pr#8121](#), Sage Weil)
- mon: remove remove\_legacy\_versions() ([pr#8324](#), Kefu Chai)
- mon: remove spurious who arg from 'mds rm ...' (John Spray)
- mon: remove unnecessary comment for update\_from\_paxos ([pr#8400](#), Qinghua Jin)
- mon: remove unused variable ([issue#15292](#), [pr#8337](#), Javier M. Mellid)
- mon: revert MonitorDBStore's WholeStoreIteratorImpl::get ([issue#13742](#), [pr#6522](#), Piotr Dałek)
- mon: should not set isvalid = true when cephx\_verify\_authorizer return false ([issue#13525](#), [pr#6306](#), Ruifeng Yang)
- mon: show the pool quota info on ceph df detail command ([issue#14216](#), [pr#7094](#), song baisen)
- mon: some cleanup in MonmapMonitor.cc ([pr#7418](#), huanwen ren)
- mon: standardize Ceph removal commands ([pr#7939](#), Dongsheng Yang)
- mon: streamline session handling, fix memory leaks (Sage Weil)
- mon: support min\_down\_reporter by subtree level (default by host) ([pr#6709](#), Xiaoxi Chen)
- mon: unconfuse object count skew message ([pr#7882](#), Piotr Dałek)
- mon: unregister command on shutdown ([pr#7504](#), huanwen ren)

- mon: upgrades must pass through hammer (Sage Weil)
- mon: warn if pg(s) not scrubbed ([issue#13142](#), [pr#6440](#), Michal Jarzabek)
- mon: warn on bogus cache tier config (Jianpeng Ma)
- mount.ceph: memory leaks ([pr#6905](#), Qiankun Zheng)
- mount.fuse.ceph: better parsing of arguments passed to mount.fuse.ceph by mount command ([issue#14735](#), [pr#7607](#), Florent Bautista)
- mrun: update path to cmake binaries ([pr#8447](#), Casey Bodley)
- msg: add override to virutal methods ([pr#6977](#), Michal Jarzabek)
- msg: add thread safety for “random” Messenger + fix wrong usage of random functions ([pr#7650](#), Avner BenHanoch)
- msg/async: AsyncConnection: avoid debug log in cleanup\_handler ([pr#7547](#), Haomai Wang)
- msg/async: AsyncMessenger: fix several bugs ([pr#7831](#), Haomai Wang)
- msg/async: AsyncMessenger: fix valgrind leak ([pr#7725](#), Haomai Wang)
- msg/async: avoid log spam on throttle ([issue#15031](#), [pr#8263](#), Kefu Chai)
- msg/async: bunch of fixes ([pr#7379](#), Piotr Dałek)
- msg/async: cleanup dead connection and misc things ([pr#7158](#), Haomai Wang)
- msg/async: don’t calculate msg header crc when not needed ([pr#7815](#), Piotr Dałek)
- msg/async: don’t use shared\_ptr to manage EventCallback ([pr#7028](#), Haomai Wang)

- msg/async: Event: fix clock skew problem ([pr#7949](#), Wei Jin)
- msg/async: fix array boundary ([pr#7451](#), Wei Jin)
- msg: async: fix perf counter description and simplify \_send\_keepalive\_or\_ack ([pr#8046](#), xie xingguo)
- msg/async: fix potential race condition ([pr#7453](#), Haomai Wang)
- msg/async: fix send closed local\_connection message problem ([pr#7255](#), Haomai Wang)
- msg/async: let receiver ack message ASAP ([pr#6478](#), Haomai Wang)
- msg/async: reduce extra tcp packet for message ack ([pr#7380](#), Haomai Wang)
- msg/async: remove experiment feature ([pr#7820](#), Haomai Wang)
- msg: async: small cleanups ([pr#7871](#), xie xingguo)
- msg/async: smarter MSG\_MORE ([pr#7625](#), Piotr Dałek)
- msg: async: start over after failing to bind a port in specified range ([issue#14928](#), [issue#13002](#), [pr#7852](#), xie xingguo)
- msg/async: support of non-block connect in async messenger ([issue#12802](#), [pr#5848](#), Jianhui Yuan)
- msg/async: \_try\_send trim already sent for outcoming\_b1 more efficient ([pr#7970](#), Yan Jun)
- msg/async: will crash if enabling async msg because of an assertion ([pr#6640](#), Zhi Zhang)
- msg: filter out lo addr when bind osd addr ([pr#7012](#), Ji Chen)
- msgr: add ceph\_perf\_msgr tool (Hoamai Wang)
- msgr: async: fix seq handling (Haomai Wang)
- msgr: async: many many fixes (Haomai Wang)
- msg: removed unneeded includes from Dispatcher ([pr#6814](#), Michal Jarzabek)
- msg: remove duplicated code - local\_delivery will now call 'enqueue' ([pr#7948](#), Avner BenHanoch)
- msg: remove unneeded inline ([pr#6989](#), Michal Jarzabek)
- msgr: fix large message data content length causing overflow ([pr#6809](#), Jun Huang, Haomai Wang)

- msgr: simple: fix clear\_pipe (#11381 Haomai Wang)
- msgr: simple: fix connect\_seq assert (Haomai Wang)
- msgr: xio: fastpath improvements (Raju Kurunkad)
- msgr: xio: fix ip and nonce (Raju Kurunkad)
- msgr: xio: improve lane assignment (Vu Pham)
- msgr: xio: misc fixes (#10735 Matt Benjamin, Kefu Chai, Danny Al-Gaaf, Raju Kurunkad, Vu Pham, Casey Bodley)
- msgr: xio: sync with accellio v1.4 (Vu Pham)
- msg: significantly reduce minimal memory usage of connections ([pr#7567](#), Piotr Dałek)
- msg/simple: pipe: memory leak when signature check failed ([pr#7096](#), Ruifeng Yang)
- msg/simple: remove unneeded friend declarations ([pr#6924](#), Michal Jarzabek)
- msg: unit tests (Haomai Wang)
- msg/xio: fix compilation ([pr#7479](#), Roi Dayan)
- msg/xio: fixes ([pr#7603](#), Roi Dayan)
- mstart: start rgw on different ports as well ([pr#8167](#), Abhishek Lekshmanan)
- nfs for rgw (Matt Benjamin, Orit Wasserman) ([pr#7634](#), Yehuda Sadeh, Matt Benjamin)
- objectcacher: misc bug fixes (Jianpeng Ma)
- objecter: avoid recursive lock of Objecter::rwlock ([pr#7343](#), Yan, Zheng)
- organizationmap: modify org mail info. ([pr#7240](#), Xiaowei Chen)
- os/bluestore: a few fixes ([pr#8193](#), Sage Weil)
- os/bluestore/BlueFS: Before reap ioct, it should wait io complete ([pr#8178](#), Jianpeng Ma)
- os/bluestore/BlueStore: Don't leak trim overlay data before write. ([pr#7895](#), Jianpeng Ma)
- os/bluestore: ceph-bluefs-tool fixes ([issue#15261](#), [pr#8292](#), Venky Shankar)
- os/bluestore: clone overlay data ([pr#7860](#), Jianpeng Ma)
- os/bluestore: fix assert ([issue#14436](#), [pr#7293](#), xie xingguo)

- os/bluestore: fix a typo in SPDK path parsing ([pr#7601](#), Jianjian Huo)
- os/bluestore: fix bluestore\_wal\_transaction\_t encoding test ([pr#7342](#), Kefu Chai)
- os/bluestore: fix bluestore\_wal\_transaction\_t encoding test ([pr#7419](#), Kefu Chai, Brad Hubbard)
- os/bluestore: insert new onode to the front position of onode LRU ([pr#7492](#), Jianjian Huo)
- os/bluestore/KernelDevice: force block size ([pr#8006](#), Sage Weil)
- os/bluestore: make bluestore\_sync\_transaction = true can work. ([pr#7674](#), Jianpeng Ma)
- os/bluestore/NVMEDevice: make IO thread using dpdk launch ([pr#8160](#), Haomai Wang)
- os/bluestore/NVMEDevice: refactor probe/attach codes and support zero command ([pr#7647](#), Haomai Wang)
- os/bluestore: revamp BlueFS bdev management and add perfcounters ([issue#15376](#), [pr#8431](#), Sage Weil)
- os/bluestore: small fixes in bluestore StupidAllocator ([pr#8101](#), Jianjian Huo)
- os/bluestore: use intrusive\_ptr for Dir ([pr#7247](#), Igor Fedotov)
- osd: add cache hint when pushing raw clone during recovery ([pr#7069](#), Zhiqiang Wang)
- osd: Add config option osd\_read\_ec\_check\_for\_errors for testing ([pr#5865](#), David Zafman)
- osd: add latency perf counters for tier operations (Xinze Chi)
- osd: add misc perfcounters (Xinze Chi)
- osd: add missing newline to usage message ([pr#7613](#), Willem Jan Withagen)
- osd: add osd op queue latency perfcounter ([pr#5793](#), Haomai Wang)
- osd: add pin/unpin support to cache tier (11066) ([pr#6326](#), Zhiqiang Wang)
- osd: add ‘proxy’ cache mode ([issue#12814](#), [pr#8210](#), Sage Weil)
- osd: add scrub persist/query API ([issue#13505](#), [pr#6898](#), Kefu Chai, Samuel Just)
- osd: add simple sleep injection in recovery (Sage Weil)
- osd: add the support of per pool scrub priority ([pr#7062](#), Zhiqiang Wang)
- osd: a fix for HeartbeatDispatcher and cleanups ([pr#7550](#), Kefu Chai)

- osd: Allow repair of history.last\_epoch\_started using config ([pr#6793](#), David Zafman)
- osd: allow SEEK\_HOLE/SEEK\_DATA for sparse read (Zhiqiang Wang)
- osd: auto repair EC pool ([issue#12754](#), [pr#6196](#), Guang Yang)
- osd: avoid calculating crush mapping for most ops ([pr#6371](#), Sage Weil)
- osd: avoid debug std::string initialization in PG::get/put ([pr#7117](#), Evgeniy Firsov)
- osd: avoid double-check for replaying and can\_checkpoint() in FileStore::\_check\_replay\_guard ([pr#6471](#), Ning Yao)
- osd: avoid duplicate op->mark\_started in ReplicatedBackend ([pr#6689](#), Jacek J. Łakis)
- osd: avoid dup omap sets for in pg metadata (Sage Weil)
- osd: avoid FORCE updating digest been overwritten by MAYBE when comparing scrub map ([pr#7051](#), Zhiqiang Wang)
- osd: avoid multiple hit set insertions (Zhiqiang Wang)
- osd: avoid osd\_op\_thread suicide because osd\_scrub\_sleep ([pr#7009](#), Jianpeng Ma)
- osd: avoid transaction append in some cases (Sage Weil)
- osd: bail out of \_committed\_osd\_maps if we are shutting down ([pr#8267](#), Samuel Just)
- osd: blockdevice: avoid implicit cast and add guard ([pr#7460](#), xie xingguo)
- osd: bluefs: fix alignment for odd page sizes ([pr#7900](#), Dan Mick)
- osd: bluestore: add 'override' to virtual functions ([pr#7886](#), Michal Jarzabek)
- osd: bluestore: allow \_dump\_onode dynamic accept log level ([pr#7995](#), Jianpeng Ma)
- osd: bluestore/blockdevice: use std::mutex et al ([pr#7568](#), Sage Weil)
- osd: bluestore: bluefs: fix several small bugs ([issue#14344](#), [issue#14343](#), [pr#7200](#), xie xingguo)
- osd: bluestore/BlueFS: initialize super block\_size earlier in mkfs ([pr#7535](#), Sage Weil)
- osd: bluestore: don't include when building without libaio ([issue#14207](#), [pr#7169](#), Mykola Golub)
- osd: bluestore: fix bluestore onode\_t attr leak ([pr#7125](#), Ning Yao)

- osd: bluestore: fix bluestore\_wal\_transaction\_t encoding test ([pr#7168](#), Kefu Chai)
- osd: bluestore: fix check for write falling within the same extent ([issue#14954](#), [pr#7892](#), Jianpeng Ma)
- osd: BlueStore: fix fsck and blockdevice read-relevant issue ([pr#7362](#), xie xingguo)
- osd: BlueStore: fix null pointer access ([issue#14561](#), [pr#7435](#), xie xingguo)
- osd: bluestore: fix several bugs ([issue#14259](#), [issue#14353](#), [issue#14260](#), [issue#14261](#), [pr#7122](#), xie xingguo)
- osd: bluestore: fix space rebalancing, collection split, buffered reads ([pr#7196](#), Sage Weil)
- osd: bluestore: for overwrite a extent, allocate new extent on min\_alloc\_size write ([pr#7996](#), Jianpeng Ma)
- osd: bluestore: improve fs-type verification and tidy up ([pr#7651](#), xie xingguo)
- osd: bluestore, kstore: fix nid overwritten logic ([issue#14407](#), [issue#14433](#), [pr#7283](#), xie xingguo)
- osd: bluestore: misc fixes ([pr#7658](#), Jianpeng Ma)
- osd: bluestore: more fixes ([pr#7130](#), Sage Weil)
- osd: BlueStore/NVMEDevice: fix compiling and fd leak ([pr#7496](#), xie xingguo)
- osd: bluestore: NVMEDevice: fix error handling ([pr#7799](#), xie xingguo)
- osd: bluestore: remove unneeded includes ([pr#7870](#), Michal Jarzabek)
- osd: bluestore: Revert NVMEDevice task ctor and refresh interface changes ([pr#7729](#), Haomai Wang)
- osd: bluestore updates, scrub fixes ([pr#8035](#), Sage Weil)
- osd: bluestore: use btree\_map for allocator ([pr#7269](#), Igor Fedotov, Sage Weil)
- osd: break PG removal into multiple iterations (#10198 Guang Yang)
- osd: cache proxy-write support (Zhiqiang Wang, Samuel Just)
- osd: cache tier: add config option for eviction check list size ([pr#6997](#), Yuan Zhou)
- osd: call on\_new\_interval on newly split child PG ([issue#13962](#), [pr#6778](#), Sage Weil)

- osd: cancel failure reports if we fail to rebind network ([pr#6278](#), Xinze Chi)
- osdc: Fix race condition with tick\_event and shutdown ([issue#14256](#), [pr#7151](#), Adam C. Emerson)
- osd: change mutex to spinlock to optimize thread context switch. ([pr#6492](#), Xiaowei Chen)
- osd: check do\_shutdown before do\_restart ([pr#6547](#), Xiaoxi Chen)
- osd: check health state before pre\_booting ([issue#14181](#), [pr#7053](#), Xiaoxi Chen)
- osd: check scrub state when handling map (Jianpeng Ma)
- osd: clarify the scrub result report ([pr#6534](#), Li Wang)
- osd/ClassHandler: only dlclose() the classes not missing ([pr#8354](#), Kefu Chai)
- osd: clean up CMPXATTR checks ([pr#5961](#), Jianpeng Ma)
- osd: clean up some constness, privateness (Kefu Chai)
- osd: clean up temp object if copy-from fails ([pr#8487](#), Sage Weil)
- osd: clean up temp object if promotion fails (Jianpeng Ma)
- osd: clear pg\_stat\_queue after stopping pgs ([issue#14212](#), [pr#7091](#), Sage Weil)
- osdc/Objecter: allow per-pool calls to op\_cancel\_writes (John Spray)
- osdc/Objecter: dout log after assign tid ([pr#8202](#), Xinze Chi)
- osdc/Objecter: fix narrow race with tid assignment ([issue#14364](#), [pr#7981](#), Sage Weil)
- osdc/Objecter: use full pgid hash in PGNLS ops ([pr#8378](#), Sage Weil)
- osd: configure promotion based on write recency (Zhiqiang Wang)
- osd: consider high/low mode when putting agent to sleep ([issue#14752](#), [pr#7631](#), Sage Weil)
- osd: constrain collections to meta and PGs (normal and temp) (Sage Weil)
- osd: correctly handle small osd\_scrub\_interval\_randomize\_ratio ([pr#7147](#), Samuel Just)
- osd: defer decoding of MOSDRepOp/MOSDRepOpReply ([pr#6503](#), Xinze Chi)
- osd: delay populating in-memory PG log hashmaps ([pr#6425](#), Piotr Dałek)
- osd: disable filestore\_xfs\_extsize by default ([issue#14397](#), [pr#7265](#), Ken Dreyer)

- osd: do not keep ref of old osdmap in pg ([issue#13990](#), [pr#7007](#), Kefu Chai)
- osd: don't do random deep scrubs for user initiated scrubs ([pr#6673](#), David Zafman)
- osd: don't send dup MMonGetOSDMap requests (Sage Weil, Kefu Chai)
- osd: don't update epoch and rollback\_info objects attrs if there is no need ([pr#6555](#), Ning Yao)
- osd: drop deprecated removal pg type ([pr#6970](#), Igor Podoski)
- osd: drop fiemap len=0 logic ([pr#7267](#), Sage Weil)
- osd: drop the interim set from load\_pgs() ([pr#6277](#), Piotr Dałek)
- osd: dump number of missing objects for each peer with pg query ([pr#6058](#), Guang Yang)
- osd: duplicated clear for peer\_missing ([pr#8315](#), Ning Yao)
- osd: EIO injection (David Zhang)
- osd: elminiate txn apend, ECSubWrite copy (Samuel Just)
- osd: enable perfcounters on sharded work queue mutexes ([pr#6455](#), Jacek J. Łakis)
- osd: ensure new osdmmaps commit before publishing them to pgs ([issue#15073](#), [pr#8096](#), Sage Weil)
- osd: erasure-code: drop entries according to LRU (Andreas-Joachim Peters)
- osd: erasure-code: fix SHEC floating point bug (#12936 Loic Dachary)
- osd: erasure-code: update to ISA-L 2.14 (Yuan Zhou)
- osd: filejournal: cleanup (David Zafman)
- osd: FileJournal: \_fdump wrongly returns if journal is currently unreadable. ([issue#13626](#), [pr#6406](#), xie xingguo)
- osd: FileJournal: fix return code of create method ([issue#14134](#), [pr#6988](#), xie xingguo)
- osd: FileJournal: reduce locking scope in write\_aio\_bh ([issue#12789](#), [pr#5670](#), Zhi Zhang)
- osd: filejournal: report journal entry count ([pr#7643](#), tianqing)
- osd: FileJournal: support batch peak and pop from writeq ([pr#6701](#), Xinze Chi)
- osd: FileStore: add a field indicate xattr only one chunk for set xattr. ([pr#6244](#), Jianpeng Ma)

- osd: FileStore: Added O\_DSYNC write scheme ([pr#7752](#), Somnath Roy)
- osd: FileStore: add error check for object\_map->sync() ([pr#7281](#), Chendi Xue)
- osd: FileStore: cleanup: remove obsolete option "filestore\_xattr\_use\_omap" ([issue#14356](#), [pr#7217](#), Vikhyat Umrao)
- osd: filestore: clone using splice (Jianpeng Ma)
- osd: FileStore: conditional collection of drive metadata ([pr#6956](#), Somnath Roy)
- osd: filestore: FALLOC\_FL\_PUNCH\_HOLE must be used with FALLOC\_FL\_KEEP\_SIZE ([pr#7768](#), xinxin shu)
- osd: filestore: fast abort if statfs encounters ENOENT ([pr#7703](#), xie xingguo)
- osd: FileStore: fix initialization order for m\_disable\_wbthrottle ([pr#8067](#), Samuel Just)
- osd: filestore: fix race condition with split vs collection\_move\_rename and long object names ([issue#14766](#), [pr#8136](#), Samuel Just)
- osd: filestore: fix recursive lock (Xinxin Shu)
- osd: filestore: fix result code overwritten for clone ([issue#14817](#), [issue#14827](#), [pr#7711](#), xie xingguo)
- osd: filestore: fix wrong scope of result code for error cases during mkfs ([issue#14814](#), [pr#7704](#), xie xingguo)
- osd: filestore: fix wrong scope of result code for error cases during mount ([issue#14815](#), [pr#7707](#), xie xingguo)
- osd: FileStore: LFNIndex: remove redundant local variable 'obj'. ([issue#13552](#), [pr#6333](#), xiexingguo)
- osd: FileStore: modify the format of colon ([pr#7333](#), Donghai Xu)
- osd: FileStore:: optimize lfn\_unlink ([pr#6649](#), Jianpeng Ma)
- osd: FileStore: potential memory leak if \_fgetattrs fails ([issue#13597](#), [pr#6377](#), xie xingguo)
- osd: FileStore: print file name before osd assert if read file failed ([pr#7111](#), Ji Chen)
- osd: FileStore: remove \_\_SWORD\_TYPE dependency ([pr#6263](#), John Coyle)
- osd: FileStore: remove unused local variable 'handle' ([pr#6381](#), xie xingguo)
- osd: filestore: restructure journal and op queue throttling ([pr#7767](#), Samuel Just)

- osd: FileStore: support multiple ondisk finish and apply finishers ([pr#6486](#), Xinze Chi, Haomai Wang)
- osd: FileStore: use pwritev instead of lseek+writev ([pr#7349](#), Haomai Wang, Tao Chang)
- osd: fix bogus scrub results when missing a clone ([issue#12738](#), [issue#12740](#), [pr#5783](#), David Zafman)
- osd: fix broken balance / localized read handling ([issue#13491](#), [pr#6364](#), Jason Dillaman)
- osd: fix bug in last\_\* PG state timestamps ([pr#6517](#), Li Wang)
- osd: fix bugs for omap ops ([pr#8230](#), Jianpeng Ma)
- osd: fix check\_for\_full (Henry Chang)
- osd: fix ClassHandler::ClassData::get\_filter() ([pr#6747](#), Yan, Zheng)
- osd: fix/clean up full map request handling ([pr#8446](#), Sage Weil)
- osd: fix debug message in OSD::is\_healthy ([pr#6226](#), Xiaoxi Chen)
- osd: fix dirty accounting in make\_writeable (Zhiqiang Wang)
- osd: fix dirtying info without correctly setting drity\_info field ([pr#8275](#), xie xingguo)
- osd: fix dump\_ops\_in\_flight races ([issue#8885](#), [pr#8044](#), David Zafman)
- osd: fix dup promotion lost op bug (Zhiqiang Wang)
- osd: fix endless repair when object is unrecoverable (Jianpeng Ma, Kefu Chai)
- osd: fix epoch check in handle\_pg\_create ([pr#8382](#), Samuel Just)
- osd: fixes for several cases where op result code was not checked or set ([issue#13566](#), [pr#6347](#), xie xingguo)
- osd: fix failure report handling during ms\_handle\_connect() ([pr#8348](#), xie xingguo)
- osd: fix FileStore::\_destroy\_collection error return code ([pr#6612](#), Ruifeng Yang)
- osd: fix forced prmootion for CALL ops ([issue#14745](#), [pr#7617](#), Sage Weil)
- osd: fix fusestore hanging during stop/quit ([issue#14786](#), [pr#7677](#), xie xingguo)
- osd: fix hitset object naming to use GMT (Kefu Chai)
- osd: fix inaccurate counter and skip over queueing an empty transaction ([pr#7754](#), xie xingguo)

- osd: fix incorrect throttle in WBThrottle ([pr#6713](#), Zhang Huan)
- osd: fix invalid list traversal in process\_copy\_chunk ([pr#7511](#), Samuel Just)
- osd: fix lack of object unblock when flush fails ([issue#14511](#), [pr#7584](#), Igor Fedotov)
- osd: fix log info ([pr#8273](#), Wei Jin)
- osd: fix misc memory leaks (Sage Weil)
- osd: fix MOSDOp encoding ([pr#6174](#), Sage Weil)
- osd: fix MOSDRepScrub reference counter in replica\_scrub ([pr#6730](#), Jie Wang)
- osd: fix negative degraded stats during backfill (Guang Yang)
- osd: fix null pointer access and race condition ([issue#14072](#), [pr#6916](#), xie xingguo)
- osd: fix osdmap dump of blacklist items (John Spray)
- osd: fix overload of '==' operator for pg\_stat\_t ([issue#14921](#), [pr#7842](#), xie xingguo)
- osd: fix peek\_queue locking in FileStore (Xinze Chi)
- osd: fix pg resurrection (#11429 Samuel Just)
- osd: fix promotion vs full cache tier (Samuel Just)
- osd: fix race condition for heartbeat\_need\_update ([issue#14387](#), [pr#7739](#), xie xingguo)
- osd: fix reactivate (check OSDSuperblock in mkfs() when we already have the superblock) ([issue#13586](#), [pr#6385](#), Vicente Cheng)
- osd: fix reference count, rare race condition etc. ([pr#8254](#), xie xingguo)
- osd: fix replay requeue when pg is still activating (#13116 Samuel Just)
- osd: fix return value from maybe\_handle\_cache\_detail() ([pr#7593](#), Igor Fedotov)
- osd: fix rollback\_info\_trimmed\_to before index() ([issue#13965](#), [pr#6801](#), Samuel Just)
- osd: fix scrub start hobject ([pr#7467](#), Sage Weil)
- osd: fix scrub stat bugs (Sage Weil, Samuel Just)
- osd: fix snap flushing from cache tier (again) (#11787 Samuel Just)
- osd: fix snap handling on promotion (#11296 Sam Just)

- osd: fix sparse-read result code checking logic ([issue#14151](#), [pr#7016](#), xie xingguo)
- osd: fix temp-clearing (David Zafman)
- osd: fix temp object removal after upgrade ([issue#13862](#), [pr#6976](#), David Zafman)
- osd: fix tick relevant issues ([pr#8369](#), xie xingguo)
- osd: fix trivial scrub bug ([pr#6533](#), Li Wang)
- osd: fix two scrub relevant issues ([pr#8462](#), xie xingguo)
- osd: fix unnecessary object promotion when deleting from cache pool ([issue#13894](#), [pr#7537](#), Igor Fedotov)
- osd: fix wip (l\_osd\_op\_wip) perf counter and remove repop\_map ([pr#7077](#), Xinze Chi)
- osd: fix wrongly placed assert and some cleanups ([pr#6766](#), xiexingguo, xie xingguo)
- osd: fix wrong return type of find\_osd\_on\_ip() ([issue#14872](#), [pr#7812](#), xie xingguo)
- osd: fix wrong use of right parenthesis in localized read logic ([pr#6566](#), Jie Wang)
- osd: force promotion for ops EC can't handle (Zhiqiang Wang)
- osd: gobject\_t: use ! instead of @ as a separator ([pr#7595](#), Sage Weil)
- osd: handle dup pg\_create that races with pg deletion ([pr#8033](#), Sage Weil)
- osd: handle log split with overlapping entries (#11358 Samuel Just)
- osd: ignore non-existent osds in unfound calc (#10976 Mykola Golub)
- osd: improve behavior on machines with large memory pages (Steve Capper)
- osd: improve temperature calculation for cache tier agent ([pr#4737](#), MingXin Liu)
- osd: include a temp namespace within each collection/pgid (Sage Weil)
- osd: increase default max open files (Owen Synge)
- osd: initialize last\_recalibrate field at construction ([pr#8071](#), xie xingguo)
- osd: init started to 0 ([issue#13206](#), [pr#6107](#), Sage Weil)
- osd: KeyValueStore: don't queue NULL context ([pr#6783](#), Haomai Wang)
- osd: KeyValueStore: fix return code of mkfs ([pr#7036](#), xie xingguo)

- osd: KeyValueStore: fix the name's typo of keyvaluestore\_default\_strip\_size ([pr#6375](#), Zhi Zhang)
- osd: KeyValueStore: fix wrongly placed assert ([issue#14176](#), [issue#14178](#), [pr#7047](#), xie xingguo)
- osd: keyvaluestore: misc fixes (Varada Kari)
- osd: kstore: fix a race condition in \_txc\_finish() ([pr#7804](#), Jianjian Huo)
- osd: kstore: latency breakdown ([pr#7850](#), James Liu)
- osd: kstore: several small fixes ([issue#14351](#), [issue#14352](#), [pr#7213](#), xie xingguo)
- osd: kstore: small fixes to kstore ([issue#14204](#), [pr#7095](#), xie xingguo)
- osd: kstore: sync up kstore with recent bluestore updates ([pr#7681](#), Jianjian Huo)
- osd: low and high speed flush modes (Mingxin Liu)
- osd: make backend and block device code a bit more generic ([pr#6759](#), Sage Weil)
- osd: make list\_missing query missing\_loc.needs\_recovery\_map ([pr#6298](#), Guang Yang)
- osd: make suicide timeouts individually configurable (Samuel Just)
- osdmap: remove unused local variables ([pr#6864](#), luo kexue)
- osdmap: rm nonused variable ([pr#8423](#), Wei Jin)
- osd: memstore: fix alignment of Page for test\_pageset ([pr#7587](#), Casey Bodley)
- osd: memstore: fix two bugs ([pr#6963](#), Casey Bodley, Sage Weil)
- osd: merge local\_t and op\_t txn to single one ([pr#6439](#), Xinze Chi)
- osd: merge multiple setattr calls into a setattrs call (Xinxin Shu)
- osd: min\_write\_recency\_for\_promote & min\_read\_recency\_for\_promote are tiering only ([pr#8081](#), huanwen ren)
- osd: misc FileStore fixes ([issue#14192](#), [issue#14188](#), [issue#14194](#), [issue#14187](#), [issue#14186](#), [pr#7059](#), xie xingguo)
- osd: misc fixes (Ning Yao, Kefu Chai, Xinze Chi, Zhiqiang Wang, Jianpeng Ma)
- osd: misc optimization for map utilization ([pr#6950](#), Ning Yao)
- osd, mon: fix exit issue ([pr#7420](#), Jiaying Ren)
- osd,mon: log leveldb and rocksdb to ceph log ([pr#6921](#), Sage Weil)
- osd: more fixes for incorrectly dirtying info; resend reply for duplicated scrub-

- reserve req ([pr#8291](#), xie xingguo)
- osd: move newest decode version of MOSDOp and MOSDOpReply to the front ([pr#6642](#), Jacek J. Łakis)
  - osd: move scrub in OpWQ (Samuel Just)
  - osd: new and delete ObjectStore::Transaction in a function is not necessary ([pr#6299](#), Ruifeng Yang)
  - osd: newstore: misc updates (including kv and os/fs stuff) ([pr#6609](#), Sage Weil)
  - osd: newstore prototype (Sage Weil)
  - osd: note down the number of missing clones ([pr#6654](#), Kefu Chai)
  - osd: ObjectStore internal API refactor (Sage Weil)
  - osd: Omap small bugs adapted ([pr#6669](#), Jianpeng Ma, David Zafman)
  - osd: optimize clone write path if object-map is enabled ([pr#6403](#), xinxin shu)
  - osd: optimize get\_object\_context ([pr#6305](#), Jianpeng Ma)
  - osd: optimize MOSDOp/do\_op/handle\_op ([pr#5211](#), Jacek J. Łakis)
  - osd: optimize scrub subset\_last\_update calculation ([pr#6518](#), Li Wang)
  - osd: optimize the session\_handle\_reset function ([issue#14182](#), [pr#7054](#), songbaisen)
  - osd: os/chain\_xattr: On linux use linux/limits.h for XATTR\_NAME\_MAX. ([pr#6343](#), John Coyle)
  - osd/OSD.cc: finish full\_map\_request every MOSDMap message. ([issue#15130](#), [pr#8147](#), Xiaoxi Chen)
  - osd/OSD: fix build\_past\_intervals\_parallel ([pr#8215](#), David Zafman)
  - osd/OSDMap: fix typo in summarize\_mapping\_stats ([pr#8088](#), Sage Weil)
  - osd: OSDMap: reset osd\_primary\_affinity shared\_ptr when deepish\_copy\_from ([issue#14686](#), [pr#7553](#), Xinze Chi)
  - osd: OSDService: Fix typo in osdmap comment ([pr#7275](#), Brad Hubbard)
  - osd: os: skip checking pg\_meta object existance in FileStore ([pr#6870](#), Ning Yao)
  - osd: partial revert of "ReplicatedPG: result code not correctly set in some cases." ([issue#13796](#), [pr#6622](#), Sage Weil)
  - osd: peer\_features includes self (David Zafman)

- osd: PG::activate(): handle unexpected cached\_removed\_snaps more gracefully ([issue#14428](#), [pr#7309](#), Alexey Sheplyakov)
- osd/PG: indicate in pg query output whether ignore\_history\_les would help ([pr#8156](#), Sage Weil)
- osd: PGLog: clean up read\_log ([pr#7092](#), Jie Wang)
- osd/PGLog: fix warning ([pr#8057](#), Sage Weil)
- osd: pg\_pool\_t: add dictionary for pool options ([issue#13077](#), [pr#6081](#), Mykola Golub)
- osd/PG: set epoch\_created and parent\_split\_bits for child pg ([issue#15426](#), [pr#8552](#), Kefu Chai)
- osd: pool size change triggers new interval (#11771 Samuel Just)
- osd: prepopulate needs\_recovery\_map when only one peer has missing (#9558 Guang Yang)
- osd: prevent osd\_recovery\_sleep from causing recovery-thread suicide ([pr#7065](#), Jianpeng Ma)
- osd: probabilistic cache tier promotion throttling ([pr#7465](#), Sage Weil)
- osd: randomize deep scrubbing ([pr#6550](#), Dan van der Ster, Herve Rousseau)
- osd: randomize scrub times (#10973 Kefu Chai)
- osd: recovery, peering fixes (#11687 Samuel Just)
- osd: reduce memory consumption of some structs ([pr#6475](#), Piotr Dałek)
- osd: reduce string use in coll\_t::calc\_str() ([pr#6505](#), Igor Podoski)
- osd: refactor scrub and digest recording (Sage Weil)
- osd: refuse first write to EC object at non-zero offset (Jianpeng Ma)
- osd: relax reply order on proxy read (#11211 Zhiqiang Wang)
- osd: release related sources when scrub is interrupted ([pr#6744](#), Jianpeng Ma)
- osd: release the message throttle when OpRequest unregistered ([issue#14248](#), [pr#7148](#), Samuel Just)
- osd: remove \_\_SWORD\_TYPE dependency ([pr#6262](#), John Coyle)
- osd: remove unused OSDMap::set\_weightf() ([issue#14369](#), [pr#7231](#), huanwen ren)
- osd: remove up\_thru\_pending field, which is never used ([pr#7991](#), xie xingguo)

- osd: reorder bool fields in PGLog struct ([pr#6279](#), Piotr Dałek)
- osd: Replace sprintf with faster implementation in eversion\_t::get\_key\_name ([pr#7121](#), Evgeniy Firsov)
- osd/ReplicatedPG: be more careful about calling publish\_stats\_to\_osd() ([issue#14962](#), [pr#8039](#), Greg Farnum)
- osd: replicatedpg: break out loop if we encounter fatal error during do\_pg\_op() ([issue#14922](#), [pr#7844](#), xie xingguo)
- osd: ReplicatedPG: clean up unused function ([pr#7211](#), Xiaowei Chen)
- osd/ReplicatedPG: clear watches on change after applying repops ([issue#15151](#), [pr#8163](#), Sage Weil)
- osd/ReplicatedPG: fix promotion recency logic ([issue#14320](#), [pr#6702](#), Sage Weil)
- osd: ReplicatedPG: remove unused local variables ([issue#13575](#), [pr#6360](#), xiexingguo)
- osd/ReplicatedPG::\_rollback\_to: update the OMAP flag ([issue#14777](#), [pr#8495](#), Samuel Just)
- osd: repop and lost-unfound overhaul ([pr#7765](#), Samuel Just)
- osd: require firefly features (David Zafman)
- osd: reset primary and up\_primary when building a new past\_interval. ([issue#13471](#), [pr#6240](#), xiexingguo)
- osd: resolve boot vs NOUP set + clear race ([pr#7483](#), Sage Weil)
- osd: scrub: do not assign value if read error ([pr#6568](#), Li Wang)
- osd/ScrubStore: remove unused function ([pr#8045](#), Kefu Chai)
- osd: set initial crush weight with more precision (Sage Weil)
- osd: several small cleanups ([pr#7055](#), xie xingguo)
- osd: SHEC no longer experimental
- osd: shut down if we flap too many times in a short period ([pr#6708](#), Xiaoxi Chen)
- osd: skip promote for writefull w/ FADVISE\_DONTNEED/NOCACHE ([pr#7010](#), Jianpeng Ma)
- osd: skip promotion for flush/evict op (Zhiqiang Wang)
- osd: slightly reduce actual size of pg\_log\_entry\_t ([pr#6690](#), Piotr Dałek)
- osd: small fixes to memstore ([issue#14228](#), [issue#14229](#), [issue#14227](#), [pr#7107](#), xie

xingguo)

- osd: stripe over small xattrs to fit in XFS's 255 byte inline limit (Sage Weil, Ning Yao)
- osd: support pool level recovery\_priority and recovery\_op\_priority ([pr#5953](#), Guang Yang)
- osd: sync object\_map on syncfs (Samuel Just)
- osd: take excl lock of op is rw (Samuel Just)
- osd: throttle evict ops (Yunchuan Wen)
- osd: try evicting after flushing is done ([pr#5630](#), Zhiqiang Wang)
- osd: upgrades must pass through hammer (Sage Weil)
- osd: use a temp object for recovery (Sage Weil)
- osd: use atomic to generate ceph\_tid ([pr#7017](#), Evgeniy Firsov)
- osd: use blkid to collection partition information (Joseph Handzik)
- osd: use optimized is\_zero in object\_stat\_sum\_t.is\_zero() ([pr#7203](#), Piotr Dałek)
- osd: use pg id (without shard) when referring the PG ([pr#6236](#), Guang Yang)
- osd: use SEEK\_HOLE / SEEK\_DATA for sparse copy (Xinxin Shu)
- osd: utime\_t, eversion\_t, osd\_stat\_sum\_t encoding optimization ([pr#6902](#), Xinze Chi)
- osd: WBThrottle cleanups (Jianpeng Ma)
- osd: WeightedPriorityQueue: move to intrusive containers ([pr#7654](#), Robert LeBlanc)
- osd: write file journal optimization ([pr#6484](#), Xinze Chi)
- osd: write journal header on clean shutdown (Xinze Chi)
- os/filestore: enlarge getxattr buffer size (Jianpeng Ma)
- os/filestore/FileJournal: set block size via config option ([pr#7628](#), Sage Weil)
- os/filestore: fix punch hole usage in \_zero ([pr#8050](#), Sage Weil)
- os/filestore: fix result handling logic of destroy\_collection ([pr#7721](#), xie xingguo)
- os/filestore: force lfn attrs to be written atomically, restructure name length limits ([pr#8496](#), Samuel Just)

- os/filestore: require offset == length == 0 for full object read; add test ([pr#7957](#), Jianpeng Ma)
- os/fs: fix io\_getevents argument ([pr#7355](#), Jingkai Yuan)
- os/fusestore: add error handling ([pr#7395](#), xie xingguo)
- os/keyvaluestore: kill KeyValueStore ([pr#7320](#), Haomai Wang)
- os/kstore: insert new onode to the front position of onode LRU ([pr#7505](#), xie xingguo)
- os/ObjectStore: add custom move operations for ObjectStore::Transaction ([pr#7303](#), Casey Bodley)
- os/ObjectStore: add noexcept to ensure move ctor is used ([pr#8421](#), Kefu Chai)
- os/ObjectStore: fix \_update\_op for split dest\_cid ([pr#8364](#), Sage Weil)
- os/ObjectStore: implement more efficient get\_encoded\_bytes() ([pr#7775](#), Piotr Dałek)
- os/ObjectStore: make device uuid probe output something friendly ([pr#8418](#), Sage Weil)
- os/ObjectStore: try\_move\_rename in transaction append and add coverage to store\_test ([issue#15205](#), [pr#8359](#), Samuel Just)
- packaging: add build dependency on python devel package ([pr#7205](#), Josh Durgin)
- packaging: make infernalis -> jewel upgrade work ([issue#15047](#), [pr#8034](#), Nathan Cutler)
- packaging: move cephfs repair tools to ceph-common ([issue#15145](#), [pr#8133](#), Boris Ranto, Ken Dreyer)
- PG: pg down state blocked by osd.x, lost osd.x cannot solve peering stuck ([issue#13531](#), [pr#6317](#), Xiaowei Chen)
- pybind: add ceph\_volume\_client interface for Manila and similar frameworks ([pr#6205](#), John Spray)
- pybind: add flock to libcephfs python bindings ([pr#7902](#), John Spray)
- pybind/cephfs: add symlink and its unit test ([pr#6323](#), Shang Ding)
- pybind: decode empty string in conf\_parse\_argv() correctly ([pr#6711](#), Josh Durgin)
- pybind: Ensure correct python flags are passed ([pr#7663](#), James Page)
- pybind: fix build failure, remove extraneous semicolon in method ([issue#14371](#), [pr#7235](#), Abhishek Lekshmanan)

- pybind: flag an RBD image as closed regardless of result code ([pr#8005](#), Jason Dillaman)
- pybind: Implementation of rados\_ioctx\_snapshot\_rollback ([pr#6878](#), Florent Manens)
- pybind/Makefile.am: Prevent race creating CYTHON\_BUILD\_DIR ([issue#15276](#), [pr#8356](#), Dan Mick)
- pybind: move cephfs to Cython ([pr#7745](#), John Spray, Mehdi Abaakouk)
- pybind: pep8 cleanups (Danny Al-Gaaf)
- pybind: port the rbd bindings to Cython ([issue#13115](#), [pr#6768](#), Hector Martin)
- pybind/rados: fix object lifetime issues and other bugs in aio ([pr#7778](#), Hector Martin)
- pybind/rados: python3 fix ([pr#8331](#), Mehdi Abaakouk)
- pybind/rados: use `__dealloc__` since `__del__` is ignored by cython ([pr#7692](#), Mehdi Abaakouk)
- pybind: remove next() on iterators ([pr#7706](#), Mehdi Abaakouk)
- pybind: replace `__del__` with `__dealloc__` for rbd ([pr#7708](#), Josh Durgin)
- pybind: support ioctx:exec ([pr#6795](#), Noah Watkins)
- pybind/test\_rbd: fix test\_create\_defaults ([issue#14279](#), [pr#7155](#), Josh Durgin)
- pybind: use correct subdir for rados install-exec rule ([pr#7684](#), Josh Durgin)
- pycephfs: many fixes for bindings (Haomai Wang)
- python binding of librados with cython ([pr#7621](#), Mehdi Abaakouk)
- python: use pip instead of python setup.py ([pr#7605](#), Loic Dachary)
- qa: add workunit to run ceph\_test\_rbd\_mirror ([pr#8221](#), Josh Durgin)
- qa: disable rbd/qemu-iotests test case 055 on RHEL/CentOSlibrbd: journal replay should honor inter-event dependencies ([issue#14385](#), [pr#7272](#), Jason Dillaman)
- qa: erasure-code benchmark plugin selection ([pr#6685](#), Loic Dachary)
- qa: fix filelock\_interrupt.py test (Yan, Zheng)
- qa: improve ceph-disk tests (Loic Dachary)
- qa: improve docker build layers (Loic Dachary)
- qa/krbd: Expunge generic/247 ([pr#6831](#), Douglas Fuller)

- qa: run-make-check.sh script (Loic Dachary)
- qa: update rest test cephfs calls ([issue#15309](#), [pr#8372](#), John Spray)
- qa: update rest test cephfs calls (part 2) ([issue#15309](#), [pr#8393](#), John Spray)
- qa/workunits/cephtool/test.sh: false positive fail on /tmp/obj1. ([pr#6837](#), Robin H. Johnson)
- qa/workunits/cephtool/test.sh: no ./ ([pr#6748](#), Sage Weil)
- qa/workunits/cephtool/test.sh: wait longer in ceph\_watch\_start() ([issue#14910](#), [pr#7861](#), Kefu Chai)
- qa/workunits: merge\_diff shouldn't attempt to use striping ([issue#14165](#), [pr#7041](#), Jason Dillaman)
- qa/workunits/rados/test.sh: capture stderr too ([pr#8004](#), Sage Weil)
- qa/workunits/rados/test.sh: test tmap\_migrate ([pr#8114](#), Sage Weil)
- qa/workunits/rbd: do not use object map during read flag testing ([pr#8104](#), Jason Dillaman)
- qa/workunits/rbd: new online maintenance op tests ([pr#8216](#), Jason Dillaman)
- qa/workunits/rbd: rbd-nbd test should use sudo for map/unmap ops ([issue#14221](#), [pr#7101](#), Jason Dillaman)
- qa/workunits/rbd: use POSIX function definition ([issue#15104](#), [pr#8068](#), Nathan Cutler)
- qa/workunits/rest/test.py: add confirmation to 'mds setmap' ([issue#14606](#), [pr#7982](#), Sage Weil)
- qa/workunits/rest/test.py: don't use newfs ([pr#8191](#), Sage Weil)
- qa/workunits/snaps: move snap tests into fs sub-directory ([pr#6496](#), Yan, Zheng)
- rados: add ceph:: namespace to bufferlist type ([pr#8059](#), Noah Watkins)
- rados: add -striper option to use libradosstriper (#10759 Sebastien Ponce)
- rados: bench: add -no-verify option to improve performance (Piotr Dalek)
- rados: bench: fix off-by-one to avoid writing past object\_size ([pr#6677](#), Tao Chang)
- rados bench: misc fixes (Dmitry Yatsushkevich)
- rados: fix bug for write bench ([pr#7851](#), James Liu)
- rados: fix error message on failed pool removal (Wido den Hollander)

- radosgw-admin: add ‘bucket check’ function to repair bucket index (Yehuda Sadeh)
- radosgw-admin: allow ([pr#8529](#), Orit Wasserman)
- radosgw-admin: Checking the legality of the parameters ([issue#13018](#), [pr#5879](#), Qiankun Zheng)
- radosgw-admin: Create -secret-key alias for -secret ([issue#5821](#), [pr#5335](#), Yuan Zhou)
- radosgw-admin: fix for ‘realm pull’ ([pr#8404](#), Casey Bodley)
- radosgw-admin: fix subuser modify output (#12286 Guce)
- radosgw-admin: metadata list user should return an empty list when user pool is empty ([issue#13596](#), [pr#6465](#), Orit Wasserman)
- radosgw-admin: ‘period commit’ supplies user-readable error messages ([pr#8264](#), Casey Bodley)
- rados: handle -snapid arg properly (Abhishek Lekshmanan)
- rados: implement rm -force option to force remove when full ([pr#6202](#), Xiaowei Chen)
- rados: improve bench buffer handling, performance (Piotr Dalek)
- rados: misc bench fixes (Dmitry Yatsushkevich)
- rados: new options for write benchmark ([pr#6340](#), Joaquim Rocha)
- rados: new pool import implementation (John Spray)
- rados: translate errno to string in CLI (#10877 Kefu Chai)
- rbd: accept map options config option (Ilya Dryomov)
- rbd: accept -user, refuse -i command-line optionals ([pr#6590](#), Ilya Dryomov)
- rbd: add disk usage tool (#7746 Jason Dillaman)
- rbd: additional validation for striping parameters ([pr#6914](#), Na Xie)
- rbd: add missing command aliases to refactored CLI ([issue#13806](#), [pr#6606](#), Jason Dillaman)
- rbd: add -object-size option, deprecate -order ([issue#12112](#), [pr#6830](#), Vikhyat Umrao)
- rbd: add pool name to disambiguate rbd admin socket commands ([pr#6904](#), wuxiangwei)
- rbd: add RBD pool mirroring configuration API + CLI ([pr#6129](#), Jason Dillaman)

- rbd: add support for mirror image promotion/demotion/resync ([pr#8138](#), Jason Dillaman)
- rbd: allow librados to prune the command-line for config overrides ([issue#15250](#), [pr#8282](#), Jason Dillaman)
- rbd: allow unmapping by spec (Ilya Dryomov)
- rbd: cli: fix arg parsing with -io-pattern (Dmitry Yatsushkevich)
- rbd: clone operation should default to image format 2 ([pr#8119](#), Jason Dillaman)
- rbd: correct an output string for merge-diff ([pr#7046](#), Kongming Wu)
- rbd: deprecate image format 1 ([pr#7841](#), Jason Dillaman)
- rbd: deprecate -new-format option (Jason Dillman)
- rbd: dynamically generated bash completion ([issue#13494](#), [pr#6316](#), Jason Dillaman)
- rbd: fix build with “-without-rbd” ([issue#14058](#), [pr#6899](#), Piotr Dałek)
- rbd: fix clone issssue ([issue#13553](#), [pr#6334](#), xinxin shu)
- rbd: fix error messages (#2862 Rajesh Nambiar)
- rbd: fixes for refactored CLI and related tests ([pr#6738](#), Ilya Dryomov)
- rbd: fix init-rbdmap CMDPARAMS ([issue#13214](#), [pr#6109](#), Sage Weil)
- rbd: fix link issues (Jason Dillaman)
- rbd: fix static initialization ordering issues ([pr#6978](#), Mykola Golub)
- rbd-fuse: image name can not include snap name ([pr#7044](#), Yongqiang He)
- rbd-fuse: implement mv operation ([pr#6938](#), wuxiangwei)
- rbd: improve CLI arg parsing, usage (Ilya Dryomov)
- rbd: journal: configuration via conf, cli, api and some fixes ([pr#6665](#), Mykola Golub)
- rbd: journal reset should disable/re-enable journaling feature ([issue#15097](#), [pr#8490](#), Jason Dillaman)
- rbd: make config changes actually apply ([pr#6520](#), Mykola Golub)
- rbdmap: add manpage ([issue#15212](#), [pr#8224](#), Nathan Cutler)
- rbdmap: systemd support ([issue#13374](#), [pr#6479](#), Boris Ranto)
- rbd: merge\_diff test should use new -object-size parameter instead of -order

([issue#14106](#), [pr#6972](#), Na Xie, Jason Dillaman)

- rbd-mirror: asok commands to get status and flush on Mirror and Replayer level ([pr#8235](#), Mykola Golub)
- rbd-mirror: enabling/disabling pool mirroring should update the mirroring directory ([issue#15217](#), [pr#8261](#), Ricardo Dias)
- rbd-mirror: fix image replay test failures ([pr#8158](#), Jason Dillaman)
- rbd-mirror: fix long termination due to 30sec wait in main loop ([pr#8185](#), Mykola Golub)
- rbd-mirror: fix missing increment for iterators ([pr#8352](#), runsis)
- rbd-mirror: ImageReplayer async start/stop ([pr#7944](#), Mykola Golub)
- rbd-mirror: ImageReplayer improvements ([pr#7759](#), Mykola Golub)
- rbd-mirror: implement ImageReplayer ([pr#7614](#), Mykola Golub)
- rbd-mirror: initial failover / fallback support ([pr#8287](#), Jason Dillaman)
- rbd-mirror: integrate with image sync state machine ([pr#8079](#), Jason Dillaman)
- rbd-mirror: make remote context respect env and argv config params ([pr#8182](#), Mykola Golub)
- rbd-mirror: minor fix-ups for initial skeleton implementation ([pr#7958](#), Mykola Golub)
- rbd-mirror: prevent enabling/disabling an image's mirroring when not in image mode ([issue#15267](#), [pr#8332](#), Ricardo Dias)
- rbd-mirror: remote to local cluster image sync ([pr#7979](#), Jason Dillaman)
- rbd-mirror: switch fsid over to mirror uuid ([issue#15238](#), [pr#8280](#), Ricardo Dias)
- rbd-mirror: use pool/image names in asok commands ([pr#8159](#), Mykola Golub)
- rbd-mirror: use the mirroring directory to detect candidate images ([issue#15142](#), [pr#8162](#), Ricardo Dias)
- rbd-mirror: workaround for intermingled lockdep singletons ([pr#8476](#), Jason Dillaman)
- rbd: must specify both of stripe-unit and stripe-count when specifying stripingv2 feature ([pr#7026](#), Donghai Xu)
- rbd-nbd: add copyright ([pr#7166](#), Li Wang)
- rbd-nbd: fix up return code handling ([pr#7215](#), Mykola Golub)

- rbd-nbd: network block device (NBD) support for RBD ([pr#6657](#), Yunchuan Wen, Li Wang)
- rbd-nbd: small improvements in logging and forking ([pr#7127](#), Mykola Golub)
- rbd: output formatter may not be closed upon error ([issue#13711](#), [pr#6706](#), xie xingguo)
- rbd: rbdmap improvements ([pr#6445](#), Boris Ranto)
- rbd: rbd order will be place in 22, when set to 0 in the config\_opt ([issue#14139](#), [issue#14047](#), [pr#6886](#), huanwen ren)
- rbd: rbd-replay-prep and rbd-replay improvements (Jason Dillaman)
- rbd: recognize queue\_depth kernel option (Ilya Dryomov)
- rbd: refactor cli command handling ([pr#5987](#), Jason Dillaman)
- rbd/run\_cli\_tests.sh: Reflect test failures ([issue#14825](#), [pr#7781](#), Zack Cerza)
- rbd: stripe unit/count set incorrectly from config ([pr#6593](#), Mykola Golub)
- rbd: striping parameters should support 64bit integers ([pr#6942](#), Na Xie)
- rbd: support for enabling/disabling mirroring on specific images ([issue#13296](#), [pr#8056](#), Ricardo Dias)
- rbd: support G and T units for CLI (Abhishek Lekshmanan)
- rbd: support negative boolean command-line optionals ([issue#13784](#), [pr#6607](#), Jason Dillaman)
- rbd: unbreak rbd map + cephx\_sign\_messages option ([pr#6583](#), Ilya Dryomov)
- rbd: update default image features ([pr#7846](#), Jason Dillaman)
- rbd: update rbd man page (Ilya Dryomov)
- rbd: update xfstests tests (Douglas Fuller)
- rbd: use default order from configuration when not specified ([pr#6965](#), Yunchuan Wen)
- rbd: use image-spec and snap-spec in help (Vikhyat Umrao, Ilya Dryomov)
- release-notes: draft v0.94.4 release notes ([pr#5907](#), Loic Dachary)
- release-notes: draft v0.94.4 release notes ([pr#6195](#), Loic Dachary)
- release-notes: draft v0.94.4 release notes ([pr#6238](#), Loic Dachary)
- release-notes: draft v0.94.6 release notes ([issue#13356](#), [pr#7689](#), Abhishek

Varshney, Loic Dachary)

- release-notes: draft v10.0.3 release notes ([pr#7592](#), Loic Dachary)
- release-notes: draft v10.0.4 release notes ([pr#7966](#), Loic Dachary)
- release-notes: draft v9.2.1 release notes ([issue#13750](#), [pr#7694](#), Abhishek Varshney)
- releases: what is merged where and when ? ([pr#8358](#), Loic Dachary)
- rest-bench: misc fixes (Shawn Chen)
- rest-bench: support https (#3968 Yuan Zhou)
- rgw: accept data only at the first time in response to a request ([pr#8084](#), sunspot)
- rgw: add a few more help options in admin interface ([pr#8410](#), Abhishek Lekshmanan)
- rgw: add a method to purge all associate keys when removing a subuser ([issue#12890](#), [pr#6002](#), Sangdi Xu)
- rgw: add a missing cap type ([pr#6774](#), Yehuda Sadeh)
- rgw: add an inspection to the field of type when assigning user caps ([pr#6051](#), Kongming Wu)
- rgw: add bucket request payment feature usage statistics integration ([issue#13834](#), [pr#6656](#), Javier M. Mellid)
- rgw: add compat header for TEMP\_FAILURE\_RETRY ([pr#6294](#), John Coyle)
- rgw: add default quota config ([pr#6400](#), Daniel Gryniewicz)
- rgw: add LifeCycle feature ([pr#6331](#), Ji Chen)
- rgw: add max multipart upload parts (#12146 Abhishek Dixit)
- rgw: add missing error code for admin op API ([pr#7037](#), Dunrong Huang)
- rgw: add missing headers to Swift container details (#10666 Ahmad Faheem, Dmytro Iurchenko)
- rgw: add stats to headers for account GET (#10684 Yuan Zhou)
- rgw: adds the radosgw-admin sync status command that gives a human readable status of the sync process at a specific zone ([pr#8030](#), Yehuda Sadeh)
- rgw: add support for caching of Keystone admin token. ([pr#7630](#), Radoslaw Zarzynski)

- rgw: add support for “end\_marker” parameter for GET on Swift account. ([issue#10682](#), [pr#4216](#), Radoslaw Zarzynski)
- rgw: add support for getting Swift’s DLO without manifest handling ([pr#6206](#), Radoslaw Zarzynski)
- rgw: add support for metadata upload during PUT on Swift container. ([pr#8002](#), Radoslaw Zarzynski)
- rgw: add support for Static Large Objects of Swift API ([issue#12886](#), [issue#13452](#), [pr#6643](#), Yehuda Sadeh, Radoslaw Zarzynski)
- rgw: add support for system requests over Swift API ([pr#7666](#), Radoslaw Zarzynski)
- rgw: add Trasnaction-Id to response (Abhishek Dixit)
- rgw: add X-Timestamp for Swift containers (#10938 Radoslaw Zarzynski)
- rgw: add zone delete to rgw-admin help ([pr#8184](#), Abhishek Lekshmanan)
- rgw: adjust error code when bucket does not exist in copy operation ([issue#14975](#), [pr#7916](#), Yehuda Sadeh)
- rgw: adjust the request\_uri to support absoluteURI of http request ([issue#12917](#), [pr#7675](#), Wenjun Huang)
- rgw: admin api for retrieving usage info (Ji Chen) ([pr#8031](#), Yehuda Sadeh, Ji Chen)
- rgw\_admin: orphans finish segfaults ([pr#6652](#), Igor Fedotov)
- rgw-admin: remove unused iterator and fix error message ([pr#8507](#), Karol Mroz)
- rgw\_admin: remove unused parent\_period arg ([pr#8411](#), Abhishek Lekshmanan)
- rgw: Allow an implicit tenant in case of Keystone ([pr#8139](#), Pete Zaitcev)
- rgw: allow authentication keystone with self signed certs ([issue#14853](#), [issue#13422](#), [pr#7777](#), Abhishek Lekshmanan)
- rgw: always check if token is expired (#11367 Anton Aksola, Riku Lehto)
- rgw: approximate AmazonS3 HostId error field. ([pr#7444](#), Robin H. Johnson)
- rgw: aws4 subdomain calling bugfix ([issue#15369](#), [pr#8472](#), Javier M. Mellid)
- rgw:bucket link now set the bucket.instance acl (bug fix) ([issue#11076](#), [pr#8037](#), Zengran Zhang)
- rgw: bucket request payment support ([issue#13427](#), [pr#6214](#), Javier M. Mellid)
- rgw: Bug fix for mtime anomalies in RadosGW and other places ([pr#7328](#), Adam C.

Emerson, Casey Bodley)

- rgw: build-related fixes ([pr#8076](#), Yehuda Sadeh, Matt Benjamin)
- rgw: calculate payload hash in RGWPutObj\_ObjStore only when necessary. ([pr#7869](#), Radoslaw Zarzynski)
- [rgw] Check return code in RGWFileHandle::write ([pr#7875](#), Brad Hubbard)
- rgw: check the return value when call fe->run() ([issue#14585](#), [pr#7457](#), wei qiaomiao)
- rgw: clarify the error message when trying to create an existed user ([pr#5938](#), Zeqiang Zhuang)
- rgw: cleanups to comments and messages ([pr#7633](#), Pete Zaitcev)
- rgw: content length ([issue#13582](#), [pr#6975](#), Yehuda Sadeh)
- rgw: conversion tool to repair broken multipart objects (#12079 Yehuda Sadeh)
- rgw: convert plain object to versioned (with null version) when removing ([issue#15243](#), [pr#8268](#), Yehuda Sadeh)
- rgw: delete default zone ([pr#7005](#), YankunLi)
- rgw: document layout of pools and objects (Pete Zaitcev)
- rgw: do not abort radowgw server when using admin op API with bad parameters ([issue#14190](#), [issue#14191](#), [pr#7063](#), Dunrong Huang)
- rgw: do not enclose bucket header in quotes (#11860 Wido den Hollander)
- rgw: do not prefetch data for HEAD requests (Guang Yang)
- rgw: do not preserve ACLs when copying object (#12370 Yehuda Sadeh)
- rgw: Do not send a Content-Type on a '304 Not Modified' response ([issue#15119](#), [pr#8253](#), Wido den Hollander)
- rgw: do not set content-type if length is 0 (#11091 Orit Wasserman)
- rgw: don't clobber bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: don't use end\_marker for namespaced object listing (#11437 Yehuda Sadeh)
- rgw: don't use rgw\_socket\_path if frontend is configured (#11160 Yehuda Sadeh)
- rgw: don't use s->bucket for metadata api path entry ([issue#14549](#), [pr#7408](#), Yehuda Sadeh)
- rgw: Drop a debugging message ([pr#7280](#), Pete Zaitcev)

- rgw: drop permissions of rgw/civetweb after startup ([issue#13600](#), [pr#8019](#), Karol Mroż)
- rgw: Drop unused usage\_exit from rgw\_admin.cc ([pr#7632](#), Pete Zaitcev)
- rgw: enforce Content-Length for POST on Swift cont/obj (#10661 Radosław Zarzynski)
- rgw: error out if frontend did not send all data (#11851 Yehuda Sadeh)
- rgw: expose the number of unhealthy workers through admin socket (Guang Yang)
- rgw: extend rgw\_extended\_http\_attrs to affect Swift accounts and containers as well ([pr#5969](#), Radosław Zarzynski)
- rgw: fail if parts not specified on multipart upload (#11435 Yehuda Sadeh)
- rgw: fcgi should include acconfig ([pr#7760](#), Abhishek Lekshmanan)
- rgw\_file: set owner uid, gid, and Unix mode on new objects ([pr#8321](#), Matt Benjamin)
- rgw: fix a glaring syntax error ([pr#6888](#), Pavan Rallabhandi)
- rgw: fix assignment of copy obj attributes (#11563 Yehuda Sadeh)
- rgw: fix a typo in error message ([pr#8434](#), Abhishek Lekshmanan)
- rgw: fix a typo in init-radosgw ([pr#6817](#), Zhi Zhang)
- rgw: fix broken stats in container listing (#11285 Radosław Zarzynski)
- rgw: fix bug in domain/subdomain splitting (Robin H. Johnson)
- rgw: fix casing of Content-Type header (Robin H. Johnson)
- rgw: fix civetweb max threads (#10243 Yehuda Sadeh)
- rgw: fix compilation warning ([pr#7160](#), Yehuda Sadeh)
- rgw: fix compiling error ([pr#8394](#), xie xingguo)
- rgw: fix Connection: header handling (#12298 Wido den Hollander)
- rgw: fix copy metadata, support X-Copied-From for swift (#10663 Radosław Zarzynski)
- rgw: fix data corruptions race condition (#11749 Wuxingyi)
- rgw: fix decoding of X-Object-Manifest from GET on Swift DLO (Radosław Zarzynski)
- rgw: fixes for per-period metadata logs ([pr#7827](#), Casey Bodley)

- rgw: fix GET on swift account when limit == 0 (#10683 Radoslaw Zarzynski)
- rgw: fix handling empty metadata items on Swift container (#11088 Radoslaw Zarzynski)
- rgw: fix JSON response when getting user quota (#12117 Wuxingyi)
- rgw: fix locator for objects starting with \_ (#11442 Yehuda Sadeh)
- rgw: fix lockdep false positive ([pr#8284](#), Yehuda Sadeh)
- rgw: fix log rotation (Wuxingyi)
- rgw: fix mdlog ([pr#8183](#), Orit Wasserman)
- rgw: fix multipart upload in retry path (#11604 Yehuda Sadeh)
- rgw: fix objects can not be displayed which object name does not cont... ([issue#12963](#), [pr#5738](#), Weijun Duan)
- rgw: fix openssl linkage ([pr#6513](#), Yehuda Sadeh)
- rgw: fix partial read issue in rgw\_admin and rgw\_tools ([pr#6761](#), Jiaying Ren)
- rgw: fix problem deleting objects begining with double underscores ([issue#15318](#), [pr#8488](#), Orit Wasserman)
- rgw: fix quota enforcement on POST (#11323 Sergey Arkhipov)
- rgw: fix reload on non Debian systems. ([pr#6482](#), Hervé Rousseau)
- rgw: fix reset\_loc (#11974 Yehuda Sadeh)
- rgw: fix response of delete expired objects ([issue#13469](#), [pr#6228](#), Yuan Zhou)
- rgw: fix return code on missing upload (#11436 Yehuda Sadeh)
- rgw: Fix subuser harder with tenants ([pr#7618](#), Pete Zaitcev)
- rgw: fix swift API returning incorrect account metadata ([issue#13140](#), [pr#6047](#), Sangdi Xu)
- rgw: fix sysvinit script
- rgw: fix sysvinit script w/ multiple instances (Sage Weil, Pavan Rallabhandi)
- rgw: fix the build failure ([pr#6927](#), Kefu Chai)
- rgw: fix typo in RGWHTTPClient::process error message ([pr#6424](#), Brad Hubbard)
- rgw: fix wrong check for parse() return ([pr#6797](#), Dunrong Huang)
- rgw: fix wrong etag calculation during POST on S3 bucket. ([issue#11241](#), [pr#6030](#),

Radoslaw Zarzynski)

- rgw: fix wrong handling of limit=0 during listing of Swift account. ([issue#14903](#), [pr#7821](#), Radoslaw Zarzynski)
- rgw: force content\_type for swift bucket stats requests (#12095 Orit Wasserman)
- rgw: force content type header on responses with no body (#11438 Orit Wasserman)
- rgw: generate Date header for civetweb (#10873 Radoslaw Zarzynski)
- rgw: generate new object tag when setting attrs (#11256 Yehuda Sadeh)
- rgw: highres time stamps ([pr#8108](#), Yehuda Sadeh, Adam C. Emerson, Matt Benjamin)
- rgw: improve content-length env var handling (#11419 Robin H. Johnson)
- rgw: improved support for swift account metadata (Radoslaw Zarzynski)
- rgw: improve error handling in S3/Keystone integration ([pr#7597](#), Radoslaw Zarzynski)
- rgw: improve handling of already removed buckets in expirer (Radoslaw Rzarzynski)
- rgw: increase verbosity level on RGWObjManifest line ([pr#7285](#), magicrobotmonkey)
- rgw: indexless ([pr#7786](#), Yehuda Sadeh)
- rgw: issue aio for first chunk before flush cached data (#11322 Guang Yang)
- rgw: Jewel nfs fixes 3 ([pr#8460](#), Matt Benjamin)
- rgw: keystone v3 ([pr#7719](#), Mark Barnes, Radoslaw Zarzynski)
- rgw: ldap fixes ([pr#8168](#), Matt Benjamin)
- rgw\_ldap: make ldap.h inclusion conditional ([pr#8500](#), Matt Benjamin)
- rgw: ldap (Matt Benjamin) ([pr#7985](#), Matt Benjamin)
- rgw: let radosgw-admin bucket stats return a standard json ([pr#7029](#), Ruifeng Yang)
- rgw: link against system openssl (instead of dlopen at runtime) ([pr#6419](#), Sage Weil)
- rgw: link civetweb with openssl (Sage, Marcus Watts) ([pr#7825](#), Marcus Watts, Sage Weil)
- rgw: link payer info to usage logging ([pr#7918](#), Yehuda Sadeh, Javier M. Mellid)
- rgw: log to /var/log/ceph instead of /var/log/radosgw

- rgw: make init script wait for radosgw to stop (#11140 Dmitry Yatsushkevich)
- rgw: make max put size configurable (#6999 Yuan Zhou)
- rgw: make quota/gc threads configurable (#11047 Guang Yang)
- rgw: make read user buckets backward compat (#10683 Radoslaw Zarzynski)
- rgw: mdlog trim add usage prompt ([pr#6059](#), Weijun Duan)
- rgw: merge manifests properly with prefix override (#11622 Yehuda Sadeh)
- rgw: modify command stucking when operating radosgw-admin metadata list user ([pr#7032](#), Peiyang Liu)
- rgw: modify documents and help infos' descriptions to the usage of option date when executing command "log show" ([pr#6080](#), Kongming Wu)
- rgw: modify the conditional statement in parse\_metadata\_key method. ([pr#5875](#), Zengran Zhang)
- rgw: move signal.h dependency from rgw\_front.h ([pr#7678](#), Matt Benjamin)
- rgw: Multipart ListPartsResult ETag quotes ([issue#15334](#), [pr#8387](#), Robin H. Johnson)
- rgw: multiple improvements regarding etag calculation for SLO/DLO of Swift API. ([pr#7764](#), Radoslaw Zarzynski)
- rgw: multiple Swift API compliance improvements for TempURL (Radoslaw Zarzynski) ([issue#14806](#), [issue#11163](#), [pr#7891](#), Radoslaw Zarzynski)
- rgw: multisite fixes ([pr#8013](#), Yehuda Sadeh)
- rgw: multitenancy support ([pr#6784](#), Yehuda Sadeh, Pete Zaitcev)
- rgw: new multisite merge ([issue#14549](#), [pr#7709](#), Yehuda Sadeh, Orit Wasserman, Casey Bodley, Daniel Gryniewicz)
- rgw: only scan for objects not in a namespace (#11984 Yehuda Sadeh)
- rgw: orphan detection tool (Yehuda Sadeh)
- rgw: Parse -subuser better ([pr#7279](#), Pete Zaitcev)
- rgw: pass in civetweb configurables (#10907 Yehuda Sadeh)
- rgw: prevent anonymous user from reading bucket with authenticated read ACL ([issue#13207](#), [pr#6057](#), root)
- rgw: radosgw-admin bucket check -fix not work ([pr#7093](#), Weijun Duan)
- rgw: rectify 202 Accepted in PUT response (#11148 Radoslaw Zarzynski)

- rgw: refuse to calculate digest when the s3 secret key is empty ([issue#13133](#), [pr#6045](#), Sangdi Xu)
- rgw: remove duplicated code in RGWRados::get\_bucket\_info() ([pr#7413](#), liyankun)
- rgw: remove extra check in RGWGetObj::execute ([issue#12352](#), [pr#5262](#), Javier M. Mellid)
- rgw: remove meta file after deleting bucket (#11149 Orit Wasserman)
- rgw: remove trailing :port from HTTP\_HOST header (Sage Weil)
- rgw: Remove unused code in PutMetadataAccount::execute ([pr#6668](#), Pete Zaitcev)
- rgw: remove unused variable in RGWPutMetadataBucket::execute ([pr#6735](#), Radoslaw Zarzynski)
- rgw: remove unused vector ([pr#7990](#), Na Xie)
- rgw: reset return code in when iterating over the bucket the objects ([issue#14826](#), [pr#7803](#), Orit Wasserman)
- rgw: retry RGWRemoteMetaLog::read\_log\_info() while master is down ([pr#8453](#), Casey Bodley)
- rgw: return 412 on bad limit when listing buckets (#11613 Yehuda Sadeh)
- rgw: Revert "rgw ldap" ([pr#8075](#), Yehuda Sadeh)
- rgw: rework X-Trans-ID header to conform with Swift API (Radoslaw Rzarzynski)
- rgw/rgw\_admin:fix bug about list and stats command ([pr#8200](#), Qiankun Zheng)
- rgw/rgw\_common.h: fix the RGWBucketInfo decoding ([pr#8154](#), Kefu Chai)
- rgw/rgw\_common.h: fix the RGWBucketInfo decoding ([pr#8165](#), Kefu Chai)
- rgw: RGWLib::env is not used so remove it ([pr#7874](#), Brad Hubbard)
- rgw/rgw\_orphan: check the return value of save\_state ([pr#7544](#), Boris Ranto)
- rgw/rgw\_resolve: fallback to res\_query when res\_nquery not implemented ([pr#6292](#), John Coyle)
- rgw: RGWZoneParams::create should not handle -EEXIST error ([pr#7927](#), Orit Wasserman)
- rgw: s3 encoding-type for get bucket (Jeff Weber)
- rgw: S3: set EncodingType in ListBucketResult ([pr#7712](#), Victor Makarov)
- rgw: send ETag, Last-Modified for swift (#11087 Radoslaw Zarzynski)

- rgw: set content length on container GET, PUT, DELETE, HEAD (#10971, #11036 Radoslaw Zarzynski)
- rgw: set max buckets per user in ceph.conf (Vikhyat Umrao)
- rgw: shard work over multiple librados instances (Pavan Rallabhandi)
- rgw: signature mismatch with escaped characters in url query portion ([issue#15358](#), [pr#8445](#), Javier M. Mellid)
- rgw: static large objects (Radoslaw Zarzynski, Yehuda Sadeh)
- rgw: store system object meta in cache when creating it ([issue#14678](#), [pr#7615](#), Yehuda Sadeh)
- rgw: support core file limit for radosgw daemon ([pr#6346](#), Guang Yang)
- rgw: support end marker on swift container GET (#10682 Radoslaw Zarzynski)
- rgw: support for aws authentication v4 (Javier M. Mellid) ([issue#10333](#), [pr#7720](#), Yehuda Sadeh, Javier M. Mellid)
- rgw: support for Swift expiration API (Radoslaw Rzarzynski, Yehuda Sadeh)
- rgw: support json format for admin policy API (Dunrong Huang) ([issue#14090](#), [pr#8036](#), Yehuda Sadeh, Dunrong Huang)
- rgw: swift: allow setting attributes with COPY (#10662 Ahmad Faheem, Dmytro Iurchenko)
- rgw: swift bulk delete (Radoslaw Zarzynski)
- rgw: swift: do not override sent content type (#12363 Orit Wasserman)
- rgw: swift: enforce Content-Type in response (#12157 Radoslaw Zarzynski)
- rgw: swift: fix account listing (#11501 Radoslaw Zarzynski)
- rgw: swift: fix metadata handling on copy (#10645 Radoslaw Zarzynski)
- rgw: swift: send Last-Modified header (#10650 Radoslaw Zarzynski)
- rgw: swift: set Content-Length for account GET (#12158 Radoslav Zarzynski)
- rgw: swift: set content-length on keystone tokens (#11473 Herv Rousseau)
- rgw: swift use Civetweb ssl can not get right url ([issue#13628](#), [pr#6408](#), Weijun Duan)
- rgw: swift versioning disabled ([pr#8066](#), Yehuda Sadeh, Radoslaw Zarzynski)
- rgw: sync fixes 3 ([pr#8170](#), Yehuda Sadeh)

- rgw: sync fixes 4 ([pr#8190](#), Yehuda Sadeh)
- rgw sync fixes ([pr#8095](#), Yehuda Sadeh)
- rgw: the map 'headers' is assigned a wrong value ([pr#8481](#), weiqiaomiao)
- rgw: try to parse Keystone token in order appropriate to configuration. ([pr#7822](#), Radoslaw Zarzynski)
- rgw: update keystone cache with token info (#11125 Yehuda Sadeh)
- rgw: update to latest civetweb, enable config for IPv6 (#10965 Yehuda Sadeh)
- rgw: use attrs from source bucket on copy (#11639 Javier M. Mellid)
- rgw: use correct oid for gc chains (#11447 Yehuda Sadeh)
- rgw:Use count fn in RGWUserBuckets for quota check ([pr#8294](#), Abhishek Lekshmanan)
- rgw: use pimpl pattern for RGWPeriodHistory ([pr#7809](#), Casey Bodley)
- rgw: user quota may not adjust on bucket removal ([issue#14507](#), [pr#7586](#), root)
- rgw: user rm is idempotent (Orit Wasserman)
- rgw: use smart pointer for C\_Reinitwatch ([pr#6767](#), Orit Wasserman)
- rgw: use unique request id for civetweb (#10295 Orit Wasserman)
- rgw: warn on suspicious civetweb frontend parameters ([pr#6944](#), Matt Benjamin)
- rocksdb: add perf counters for get/put latency (Xinxin Shu)
- rocksdb: build with PORTABLE=1 ([pr#6311](#), Sage Weil)
- rocksdb, leveldb: fix compact\_on\_mount (Xiaoxi Chen)
- rocksdb: pass options as single string (Xiaoxi Chen)
- rocksdb: remove rdb source files from dist tarball ([issue#13554](#), [pr#6379](#), Kefu Chai)
- rocksdb: remove rdb sources from dist tarball ([issue#13554](#), [pr#7105](#), Venky Shankar)
- rocksdb: update to latest (Xiaoxi Chen)
- rocksdb: use native rocksdb makefile (and our autotools) ([pr#6290](#), Sage Weil)
- rpm: add suse firewall files (Tim Serong)
- rpm: always rebuild and install man pages for rpm (Owen Synge)
- rpm: ceph.spec.in: correctly declare systemd dependency for SLE/openSUSE

([pr#6114](#), Nathan Cutler)

- rpm: ceph.spec.in: fix libs-compat / devel-compat conditional ([issue#12315](#), [pr#5219](#), Ken Dreyer)
- rpm,deb: remove conditional BuildRequires for btrfs-progs ([issue#15042](#), [pr#8016](#), Erwan Velu)
- rpm: loosen ceph-test dependencies (Ken Dreyer)
- rpm: many spec file fixes (Owen Synge, Ken Dreyer)
- rpm: misc fixes (Boris Ranto, Owen Synge, Ken Dreyer, Ira Cooper)
- rpm: misc systemd and SUSE fixes (Owen Synge, Nathan Cutler)
- rpm: move %post(un) ldconfig calls to ceph-base ([issue#14940](#), [pr#7867](#), Nathan Cutler)
- rpm: move runtime dependencies to ceph-base and fix other packaging issues ([issue#14864](#), [pr#7826](#), Nathan Cutler)
- rpm: prefer UID/GID 167 when creating ceph user/group ([issue#15246](#), [pr#8277](#), Nathan Cutler)
- rpm: remove sub-package dependencies on “ceph” ([issue#15146](#), [pr#8137](#), Ken Dreyer)
- rpm: rhel 5.9 librados compile fix, moved blkid to RBD check/compilation ([issue#13177](#), [pr#5954](#), Rohan Mars)
- script: add missing stop\_rgw variable to stop.sh script ([pr#7959](#), Karol Mroz)
- scripts: adjust mstart and mstop script to run with cmake build ([pr#6920](#), Orit Wasserman)
- scripts: release\_notes can track original issue ([pr#6009](#), Abhishek Lekshmanan)
- script: subscription-manager support ([issue#14972](#), [pr#7907](#), Loic Dachary)
- selinux: allow log files to be located in /var/log/radosgw ([pr#7604](#), Boris Ranto)
- selinux policy (Boris Ranto, Milan Broz)
- selinux: Update policy to grant additional access ([issue#14870](#), [pr#7971](#), Boris Ranto)
- set 128MB tcmalloc cache size by bytes ([pr#8427](#), Star Guo)
- sstring.hh: return type from str\_len(...) need not be const ([pr#7679](#), Matt Benjamin)
- stringify outputted error code and fix unmatched parentheses. ([pr#6998](#),

- xie.xingguo, xie xingguo)
- Striper: reduce assemble\_result log level ([pr#8426](#), Jason Dillaman)
- submodules: revert an accidental change ([pr#7929](#), Yehuda Sadeh)
- systemd: correctly escape block device paths ([issue#14706](#), [pr#7579](#), James Page)
- systemd: drop any systemd imposed process/thread limits ([pr#8450](#), James Page)
- systemd: fix typos ([pr#6679](#), Tobias Suckow)
- systemd: logrotate fixes (Tim Serong, Lars Marowsky-Bree, Nathan Cutler)
- systemd: many fixes (Sage Weil, Owen Synge, Boris Ranto, Dan van der Ster)
- systemd: run daemons as user ceph
- systemd: set up environment in rbdmap unit file ([issue#14984](#), [pr#8222](#), Nathan Cutler)
- systemd: start/stop/restart ceph services by daemon type ([issue#13497](#), [pr#6276](#), Zhi Zhang)
- sysvinit: allow custom cluster names ([pr#6732](#), Richard Chan)
- sysvinit compat: misc fixes (Owen Synge)
- test: add missing shut\_down mock method ([pr#8125](#), Jason Dillaman)
- test/bufferlist: Avoid false-positive tests ([pr#7955](#), Erwan Velu)
- test: ceph\_test\_rados: use less CPU ([pr#7513](#), Samuel Just)
- test/cli-integration/rbd: disable progress output ([issue#14931](#), [pr#7858](#), Josh Durgin)
- test: correct librbd errors discovered with unoptimized cmake build ([pr#7914](#), Jason Dillaman)
- test: create pools for rbd tests with different prefix ([pr#7738](#), Mykola Golub)
- test: enable test for bug #2339 which has been resolved. ([pr#7743](#), You Ji)
- test/encoding/readable.sh fix ([pr#6714](#), Igor Podoski)
- Test exit values on test.sh, fix tier.cc ([issue#15165](#), [pr#8266](#), Samuel Just)
- test: fix issues discovered via the rbd permissions test case ([pr#8129](#), Jason Dillaman)
- test: fix osd-scrub-snaps.sh ([pr#6697](#), Xinze Chi)

- test: Fix test to run with btrfs which has snap\_### dirs ([issue#15347](#), [pr#8420](#), David Zafman)
- test: fixup and improvements for rbd-mirror test ([pr#8090](#), Mykola Golub)
- test: fix ut test failure caused by lfn change ([issue#15464](#), [pr#8544](#), xie xingguo)
- test: fix valgrind memcheck issues for rbd-mirror test cases ([issue#15354](#), [pr#8493](#), Jason Dillaman)
- test: handle exception thrown from close during rbd lock test ([pr#8124](#), Jason Dillaman)
- test/libcephfs/flock: add sys/file.h include for flock operations ([pr#6310](#), John Coyle)
- test/librados/test.cc: clean up EC pools' crush rules too ([issue#13878](#), [pr#6788](#), Loic Dachary, Dan Mick)
- test/librbd/fsx: Use c++11 std::mt19937 generator instead of random\_r() ([pr#6332](#), John Coyle)
- test: misc fs test improvements (John Spray, Loic Dachary)
- test/mon/osd-erasure-code-profile: pick new mon port ([pr#7161](#), Sage Weil)
- test: more debug logging for TestWatchNotify ([pr#7737](#), Mykola Golub)
- test: new librbd flatten test case ([pr#7609](#), Jason Dillaman)
- test/osd: Relax the timing intervals in osd-markdown.sh ([pr#7899](#), Dan Mick)
- test\_pool\_create.sh: put test files in the test dir so they are cleaned up ([pr#8219](#), Josh Durgin)
- test/pybind/test\_ceph\_argparse: fix reweight-by-utilization tests ([pr#8027](#), Kefu Chai, Sage Weil)
- test: python tests, linter cleanup (Alfredo Deza)
- test/radosgw-admin: update the expected usage outputs ([pr#7723](#), Kefu Chai)
- test: rbd-mirror: add "switch to the next tag" test ([pr#8149](#), Mykola Golub)
- test: rbd-mirror: compare positions using all fields ([pr#8172](#), Mykola Golub)
- test: rbd-mirror: script improvements for manual testing ([pr#8325](#), Mykola Golub)
- test: reproducer for writeback Cow deadlock ([pr#8009](#), Jason Dillaman)
- test/rgw: add multisite test for meta sync across periods ([pr#7887](#), Casey Bodley)

- test\_rgw\_admin: use freopen for output redirection. ([pr#6303](#), John Coyle)
- tests: add const for ec test ([pr#6911](#), Michal Jarzabek)
- tests: add Ubuntu 16.04 xenial dockerfile ([pr#8519](#), Loic Dachary)
- tests: allow docker-test.sh to run under root ([issue#13355](#), [pr#6173](#), Loic Dachary)
- tests: allow object corpus readable test to skip specific incompat instances ([pr#6932](#), Igor Podoski)
- tests: centos7 needs the Continuous Release (CR) Repository enabled for ([issue#13997](#), [pr#6842](#), Brad Hubbard)
- tests: ceph-disk.sh: should use “readlink -f” instead ([pr#7594](#), Kefu Chai)
- tests: ceph-disk.sh: use “readlink -f” instead for fullpath ([pr#7606](#), Kefu Chai)
- tests: ceph-disk workunit uses configobj ([pr#6342](#), Loic Dachary)
- tests: ceph-helpers assert success getting backfills ([pr#6699](#), Loic Dachary)
- tests: ceph\_test\_keyvaluedb\_iterators: fix broken test ([pr#6597](#), Haomai Wang)
- tests: concatenate test\_rados\_test\_tool from src and qa ([issue#13691](#), [pr#6464](#), Loic Dachary)
- tests: configure with rocksdb by default ([issue#14220](#), [pr#7100](#), Loic Dachary)
- tests: destroy testprofile before creating one ([issue#13664](#), [pr#6446](#), Loic Dachary)
- tests: fix a few build warnings ([pr#7608](#), Sage Weil)
- tests: fixes for rbd xstests (Douglas Fuller)
- tests: fix failure for osd-scrub-snap.sh ([issue#13986](#), [pr#6890](#), Loic Dachary, Ning Yao)
- tests: Fix for make check. ([pr#7102](#), David Zafman)
- tests: Fixing broken test/cephtool-test-mon.sh test ([pr#8429](#), Erwan Velu)
- tests: fix race condition testing auto scrub ([issue#13592](#), [pr#6724](#), Xinze Chi, Loic Dachary)
- tests: fix test\_rados\_tools.sh rados lookup ([issue#13691](#), [pr#6502](#), Loic Dachary)
- tests: fix tiering health checks (Loic Dachary)
- tests: fix typo in TestClsRbd.snapshots test case ([issue#13727](#), [pr#6504](#), Jason Dillaman)

- tests: flush op work queue prior to destroying MockImageCtx ([issue#14092](#), [pr#7002](#), Jason Dillaman)
- tests for low-level performance (Haomai Wang)
- tests: ignore test-suite.log ([pr#6584](#), Loic Dachary)
- tests: Improving 'make check' execution time ([pr#8131](#), Erwan Velu)
- tests: many ec non-regression improvements (Loic Dachary)
- tests: many many ec test improvements (Loic Dachary)
- tests: notification slave needs to wait for master ([issue#13810](#), [pr#7220](#), Jason Dillaman)
- tests: -osd-scrub-load-threshold=2000 for more consistency ([issue#14027](#), [pr#6871](#), Loic Dachary)
- tests: osd-scrub-snaps.sh to display full osd logs on error ([issue#13986](#), [pr#6857](#), Loic Dachary)
- tests: port uniqueness reminder ([pr#6387](#), Loic Dachary)
- tests: restore run-cli-tests ([pr#6571](#), Loic Dachary, Sage Weil, Jason Dillaman)
- tests: snap rename and rebuild object map in client update test ([pr#7224](#), Jason Dillaman)
- tests: sync ceph-erasure-code-corpus for mktemp -d ([pr#7596](#), Loic Dachary)
- tests: test/librados/test.cc must create profile ([issue#13664](#), [pr#6452](#), Loic Dachary)
- tests: test\_pidfile.sh lingering processes ([issue#14834](#), [pr#7734](#), Loic Dachary)
- tests: unittest\_bufferlist: fix hexdump test ([pr#7152](#), Sage Weil)
- tests: unittest\_ipaddr: fix segv ([pr#7154](#), Sage Weil)
- test/system/rados\_list\_parallel: print oid if rados\_write fails ([issue#15240](#), [pr#8309](#), Kefu Chai)
- test/system/\*: use dynamically generated pool name ([issue#15240](#), [pr#8318](#), Kefu Chai)
- test/test-erasure-code.sh: disable pg\_temp priming ([issue#15211](#), [pr#8260](#), Sage Weil)
- test: TestMirroringWatcher test cases were not closing images ([pr#8435](#), Jason Dillaman)

- test/TestPGLog: fix the FTBFS ([issue#14930](#), [pr#7855](#), Kefu Chai)
- test/test\_pool\_create.sh: fix port ([pr#8361](#), Sage Weil)
- test/time: no need to abs(uint64\_t) for comparing ([pr#7726](#), Kefu Chai)
- test: update rbd integration cram tests for new default features ([pr#8001](#), Jason Dillaman)
- test: use sequential journal\_tid for object cacher test ([issue#13877](#), [pr#6710](#), Josh Durgin)
- tools: add cephfs-table-tool ‘take\_inos’ ([pr#6655](#), John Spray)
- tools/cephfs: add tmap\_upgrade ([pr#7003](#), John Spray)
- tools/cephfs: fix overflow writing header to fixed size buffer (#13816) ([pr#6617](#), John Spray)
- tools/cephfs: fix tmap\_upgrade ([issue#15135](#), [pr#8128](#), John Spray)
- tools: ceph\_monstore\_tool: add inflate-pgmap command ([issue#14217](#), [pr#7097](#), Kefu Chai)
- tools: ceph-monstore-update-crush: add “-test” when testing crushmap ([pr#6418](#), Kefu Chai)
- tools: Fix layout handing in cephfs-data-scan (#13898) ([pr#6719](#), John Spray)
- tools: monstore: add ‘show-versions’ command. ([pr#7073](#), Cilang Zhao)
- tools/rados: reduce “rados put” memory usage by op\_size ([pr#7928](#), Piotr Dałek)
- tools:remove duplicate references ([pr#5917](#), Bo Cai)
- tools: support printing part cluster map in readable fashion ([issue#13079](#), [pr#5921](#), Bo Cai)
- unittest\_compression\_zlib: do not assume buffer will be null terminated ([pr#8064](#), Sage Weil)
- unittest\_erasure\_code\_plugin: fix deadlock (Alpine) ([pr#8314](#), John Coyle)
- unittest\_osdmap: default crush tunables now firefly ([pr#8098](#), Sage Weil)
- upstart: throttle restarts (#11798 Sage Weil, Greg Farnum)
- vstart: fix up cmake paths when VSTART\_DEST is given ([pr#8363](#), Casey Bodley)
- vstart: grant full access to Swift testing account ([pr#6239](#), Yuan Zhou)
- vstart: make -k with optional mon\_num. ([pr#8251](#), Jianpeng Ma)

- vstart: set cephfs root uid/gid to caller ([pr#6255](#), John Spray)
- vstart.sh: add mstart, mstop, mrun wrappers for running multiple vstart-style test clusters out of src tree ([pr#6901](#), Yehuda Sadeh)
- vstart.sh: avoid race condition starting rgw via vstart.sh ([issue#14829](#), [pr#7727](#), Javier M. Mellid)
- vstart.sh: silence a harmless msg where btrfs is not found ([pr#7640](#), Patrick Donnelly)
- xio: add prefix to xio msgr logs ([pr#8148](#), Roi Dayan)
- xio: fix compilation against latest accelio ([pr#8022](#), Roi Dayan)
- xio: fix incorrect ip being assigned in case of multiple RDMA ports ([pr#7747](#), Subramanyam Varanasi)
- xio: remove duplicate assignment of peer addr ([pr#8025](#), Roi Dayan)
- xio: remove redundant magic methods ([pr#7773](#), Roi Dayan)
- xio: remove unused variable ([pr#8023](#), Roi Dayan)
- xio: xio\_init needs to be called before any other xio function ([pr#8227](#), Roi Dayan)
- xxhash: use clone of xxhash.git; add .gitignore ([pr#7986](#), Sage Weil)

## v9.2.1 Infernalis

This Infernalis point release fixes several packagins and init script issues, enables the librbd objectmap feature by default, a few librbd bugs, and a range of miscellaneous bug fixes across the system.

We recommend that all infernalis v9.2.0 users upgrade.

For more detailed information, see [the complete changelog](#).

## Upgrading

- Some symbols wrongly exposed by the C++ interface for librados in v9.1.0 and v9.2.0 were removed. If you compiled your own application against librados shipped with these releases, it is very likely referencing these removed symbols. So you will need to recompile it.

## Notable Changes

- build/ops: Ceph daemon failed to start, because the service name was already used. ([issue#13474](#), [pr#6833](#), Chuanhong Wang)
- build/ops: ceph upstart script rbdmap.conf incorrectly processes parameters ([issue#13214](#), [pr#6396](#), Sage Weil)
- build/ops: libunwind package missing on CentOS 7 ([issue#13997](#), [pr#6845](#), Loic Dachary)
- build/ops: rbd-replay-\* moved from ceph-test-dbg to ceph-common-dbg as well ([issue#13785](#), [pr#6628](#), Loic Dachary)
- build/ops: systemd/ceph-disk@.service assumes /bin/flock ([issue#13975](#), [pr#6852](#), Loic Dachary)
- build/ops: systemd: no rbdmap systemd unit file ([issue#13374](#), [pr#6500](#), Boris Ranto)
- common: auth/cephx: large amounts of log are produced by osd ([issue#13610](#), [pr#6836](#), Qiankun Zheng)
- common: log: Log.cc: Assign LOG\_DEBUG priority to syslog calls ([issue#13993](#), [pr#6993](#), Brad Hubbard)
- crush: crash if we see CRUSH\_ITEM\_NONE in early rule step ([issue#13477](#), [pr#6626](#), Sage Weil)
- fs: Ceph file system is not freeing space ([issue#13777](#), [pr#7431](#), Yan, Zheng, John)

Spray)

- fs: Ceph-fuse won't start correctly when the option log\_max\_new in ceph.conf set to zero ([issue#13443](#), [pr#6395](#), Wenjun Huang)
- fs: Segmentation fault accessing file using fuse mount ([issue#13714](#), [pr#6853](#), Yan, Zheng)
- librbd: Avoid re-writing old-format image header on resize ([issue#13674](#), [pr#6630](#), Jason Dillaman)
- librbd: ImageWatcher shouldn't block the notification thread ([issue#14373](#), [pr#7406](#), Jason Dillaman)
- librbd: QEMU hangs after creating snapshot and stopping VM ([issue#13726](#), [pr#6632](#), Jason Dillaman)
- librbd: Verify self-managed snapshot functionality on image create ([issue#13633](#), [pr#7080](#), Jason Dillaman)
- librbd: [ FAILED ] TestLibRBD.SnapRemoveViaLockOwner ([issue#14164](#), [pr#7079](#), Jason Dillaman)
- librbd: enable feature objectmap ([issue#13558](#), [pr#6477](#), xinxin shu)
- librbd: fix merge-diff for >2GB diff-files ([issue#14030](#), [pr#6981](#), Jason Dillaman)
- librbd: flattening an rbd image with active IO can lead to hang ([issue#14092](#), [issue#14483](#), [pr#7484](#), Jason Dillaman)
- mds: fix client capabilities during reconnect (client.XXXX isn't responding to mclientcaps warning) ([issue#11482](#), [pr#6752](#), Yan, Zheng)
- mon: Ceph Pools' MAX AVAIL is 0 if some OSDs' weight is 0 ([issue#13840](#), [pr#6907](#), Chengyuan Li)
- mon: should not set isvalid = true when cephx\_verify\_authorizer return... ([issue#13525](#), [pr#6392](#), Ruifeng Yang)
- objecter: pool op callback may hang forever. ([issue#13642](#), [pr#6627](#), xie xingguo)
- objecter: potential null pointer access when do pool\_snap\_list. ([issue#13639](#), [pr#6840](#), xie xingguo)
- osd: FileStore: potential memory leak if getattr fails. ([issue#13597](#), [pr#6846](#), xie xingguo)
- osd: OSD::build\_past\_intervals\_parallel() shall reset primary and up\_primary when begin a new past\_interval. ([issue#13471](#), [pr#6397](#), xiexinguo)
- osd: call on\_new\_interval on newly split child PG ([issue#13962](#), [pr#6849](#), Sage Weil)

- osd: ceph-disk list fails on /dev/cciss!c0d0 ([issue#13970](#), [issue#14230](#), [pr#6880](#), Loic Dachary)
- osd: ceph-disk: use blkid instead of sgdisk -i ([issue#14080](#), [pr#7001](#), Loic Dachary, Ilya Dryomov)
- osd: fix race condition during send\_failures ([issue#13821](#), [pr#6694](#), Sage Weil)
- osd: osd/PG.cc: 288: FAILED assert(info.last\_epoch\_started >= info.history.last\_epoch\_started) ([issue#14015](#), [pr#6851](#), David Zafman)
- osd: pgs stuck inconsistent after infernalis upgrade ([issue#13862](#), [pr#7421](#), David Zafman)
- rbd: TaskFinisher::cancel should remove event from SafeTimer ([issue#14476](#), [pr#7426](#), Douglas Fuller)
- rbd: cls\_rbd: object\_map\_save should enable checksums ([issue#14280](#), [pr#7428](#), Douglas Fuller)
- rbd: misdirected op in rbd balance-reads test ([issue#13491](#), [pr#6629](#), Jason Dillaman)
- rbd: pure virtual method called ([issue#13636](#), [pr#6633](#), Jason Dillaman)
- rbd: rbd clone issue ([issue#13553](#), [pr#6474](#), xinxin shu)
- rbd: rbd-replay does not check for EOF and goes to endless loop ([issue#14452](#), [pr#7427](#), Mykola Golub)
- rbd: unknown argument -quiet in udevadm settle ([issue#13560](#), [pr#6634](#), Jason Dillaman)
- rgw: init script reload doesn't work on EL7 ([issue#13709](#), [pr#6650](#), Hervé Rousseau)
- rgw: radosgw-admin -help doesn't show the orphans find command ([issue#14516](#), [pr#7543](#), Yehuda Sadeh)
- tests: ceph-disk workunit uses configobj ([issue#14004](#), [pr#6828](#), Loic Dachary)
- tests: fsx failed to compile ([issue#14384](#), [pr#7429](#), Greg Farnum)
- tests: notification slave needs to wait for master ([issue#13810](#), [pr#7225](#), Jason Dillaman)
- tests: rebuild exclusive lock test should acquire exclusive lock ([issue#14121](#), [pr#7038](#), Jason Dillaman)
- tests: testprofile must be removed before it is re-created ([issue#13664](#), [pr#6449](#), Loic Dachary)

- tests: verify it is possible to reuse an OSD id ([issue#13988](#), [pr#6882](#), Loic Dachary)

## v9.2.0 Infernalis

---

This major release will be the foundation for the next stable series. There have been some major changes since v0.94.x Hammer, and the upgrade process is non-trivial. Please read these release notes carefully.

### Major Changes from Hammer

---

- *General:*
  - Ceph daemons are now managed via systemd (with the exception of Ubuntu Trusty, which still uses upstart).
  - Ceph daemons run as ‘ceph’ user instead root.
  - On Red Hat distros, there is also an SELinux policy.
- *RADOS:*
  - The RADOS cache tier can now proxy write operations to the base tier, allowing writes to be handled without forcing migration of an object into the cache.
  - The SHEC erasure coding support is no longer flagged as experimental. SHEC trades some additional storage space for faster repair.
  - There is now a unified queue (and thus prioritization) of client IO, recovery, scrubbing, and snapshot trimming.
  - There have been many improvements to low-level repair tooling (ceph-objectstore-tool).
  - The internal ObjectStore API has been significantly cleaned up in order to facilitate new storage backends like NewStore.
- *RGW:*
  - The Swift API now supports object expiration.
  - There are many Swift API compatibility improvements.
- *RBD:*
  - The `rbd du` command shows actual usage (quickly, when object-map is enabled).

- The object-map feature has seen many stability improvements.
  - Object-map and exclusive-lock features can be enabled or disabled dynamically.
  - You can now store user metadata and set persistent librbd options associated with individual images.
  - The new deep-flatten features allows flattening of a clone and all of its snapshots. (Previously snapshots could not be flattened.)
  - The export-diff command command is now faster (it uses aio). There is also a new fast-diff feature.
  - The --size argument can be specified with a suffix for units (e.g., `--size 64G`).
  - There is a new `rbd status` command that, for now, shows who has the image open/mapped.
- *CephFS*:
    - You can now rename snapshots.
    - There have been ongoing improvements around administration, diagnostics, and the check and repair tools.
    - The caching and revocation of client cache state due to unused inodes has been dramatically improved.
    - The ceph-fuse client behaves better on 32-bit hosts.

## Distro compatibility

---

We have decided to drop support for many older distributions so that we can move to a newer compiler toolchain (e.g., C++11). Although it is still possible to build Ceph on older distributions by installing backported development tools, we are not building and publishing release packages for ceph.com.

We now build packages for:

- CentOS 7 or later. We have dropped support for CentOS 6 (and other RHEL 6 derivatives, like Scientific Linux 6).
- Debian Jessie 8.x or later. Debian Wheezy 7.x's g++ has incomplete support for C++11 (and no systemd).
- Ubuntu Trusty 14.04 or later. Ubuntu Precise 12.04 is no longer supported.
- Fedora 22 or later.

# Upgrading from Firefly

Upgrading directly from Firefly v0.80.z is not recommended. It is possible to do a direct upgrade, but not without downtime. We recommend that clusters are first upgraded to Hammer v0.94.4 or a later v0.94.z release; only then is it possible to upgrade to Infernalis 9.2.z for an online upgrade (see below).

To do an offline upgrade directly from Firefly, all Firefly OSDs must be stopped and marked down before any Infernalis OSDs will be allowed to start up. This fencing is enforced by the Infernalis monitor, so use an upgrade procedure like:

1. Upgrade Ceph on monitor hosts
2. Restart all ceph-mon daemons
3. Upgrade Ceph on all OSD hosts
4. Stop all ceph-osd daemons
5. Mark all OSDs down with something like:

```
1. ceph osd down `seq 0 1000`
```

6. Start all ceph-osd daemons
7. Upgrade and restart remaining daemons (ceph-mds, radosgw)

# Upgrading from Hammer

- All cluster nodes must first upgrade to Hammer v0.94.4 or a later v0.94.z release; only then is it possible to upgrade to Infernalis 9.2.z.
- For all distributions that support systemd (CentOS 7, Fedora, Debian Jessie 8.x, OpenSUSE), ceph daemons are now managed using native systemd files instead of the legacy sysvinit scripts. For example:

```
1. systemctl start ceph.target      # start all daemons
2. systemctl status ceph-osd@12    # check status of osd.12
```

The main notable distro that is *not* yet using systemd is Ubuntu trusty 14.04. (The next Ubuntu LTS, 16.04, will use systemd instead of upstart.)

- Ceph daemons now run as user and group `ceph` by default. The ceph user has a static UID assigned by Fedora and Debian (also used by derivative distributions like RHEL/CentOS and Ubuntu). On SUSE the ceph user will currently get a dynamically assigned UID when the user is created.

If your systems already have a ceph user, upgrading the package will cause

problems. We suggest you first remove or rename the existing ‘ceph’ user and ‘ceph’ group before upgrading.

When upgrading, administrators have two options:

i. Add the following line to `ceph.conf` on all hosts:

```
1. setuser match path = /var/lib/ceph/$type/$cluster-$id
```

This will make the Ceph daemons run as root (i.e., not drop privileges and switch to user ceph) if the daemon’s data directory is still owned by root. Newly deployed daemons will be created with data owned by user ceph and will run with reduced privileges, but upgraded daemons will continue to run as root.

ii. Fix the data ownership during the upgrade. This is the preferred option, but it is more work and can be very time consuming. The process for each host is to:

iii. Upgrade the ceph package. This creates the ceph user and group. For example:

```
1. ceph-deploy install --stable infernalis HOST
```

a. Stop the daemon(s):

```
1. service ceph stop          # fedora, centos, rhel, debian  
2. stop ceph-all             # ubuntu
```

b. Fix the ownership:

```
1. chown -R ceph:ceph /var/lib/ceph  
2. chown -R ceph:ceph /var/log/ceph
```

c. Restart the daemon(s):

```
1. start ceph-all           # ubuntu  
2. systemctl start ceph.target # debian, centos, fedora, rhel
```

*Alternatively*, the same process can be done with a single daemon type, for example by stopping

1. only monitors and chowning only `/var/lib/ceph/mon`.

- The on-disk format for the experimental KeyValueStore OSD backend has changed. You will need to remove any OSDs using that backend before you upgrade any test clusters that use it.
- When a pool quota is reached, librados operations now block indefinitely, the same way they do when the cluster fills up. (Previously they would return -ENOSPC). By default, a full cluster or pool will now block. If your librados application can handle ENOSPC or EDQUOT errors gracefully, you can get error returns instead by using the new librados OPERATION\_FULL\_TRY flag.

- The return code for librbd's `rbd_aio_read` and `Image::aio_read` API methods no longer returns the number of bytes read upon success. Instead, it returns 0 upon success and a negative value upon failure.
- ‘ceph scrub’, ‘ceph compact’ and ‘ceph sync force’ are now DEPRECATED. Users should instead use ‘ceph mon scrub’, ‘ceph mon compact’ and ‘ceph mon sync force’.
- ‘ceph mon\_metadata’ should now be used as ‘ceph mon metadata’. There is no need to deprecate this command (same major release since it was first introduced).
- The `-dump-json` option of “osdmaptool” is replaced by `-dump json`.
- The commands of “`pg ls-by-{pool,primary,osd}`” and “`pg ls`” now take “recovering” instead of “recovery”, to include the recovering pgs in the listed pgs.

## Notable Changes since Hammer

---

- aarch64: add optimized version of `crc32c` (Yazen Ghannam, Steve Capper)
- auth: cache/reuse crypto lib key objects, optimize msg signature check (Sage Weil)
- auth: reinit NSS after `fork()` (#11128 Yan, Zheng)
- autotools: fix out of tree build (Krzysztof Kosinski)
- autotools: improve make check output (Loic Dachary)
- buffer: add `invalidate_crc()` (Piotr Dalek)
- buffer: fix zero bug (#12252 Haomai Wang)
- buffer: some cleanup (Michal Jarzabek)
- build: allow `tcmalloc-minimal` (Thorsten Behrens)
- build: C++11 now supported
- build: cmake: fix nss linking (Danny Al-Gaaf)
- build: cmake: misc fixes (Orit Wasserman, Casey Bodley)
- build: disable LTTNG by default (#11333 Josh Durgin)
- build: do not build ceph-dencoder with `tcmalloc` (#10691 Boris Ranto)
- build: fix junit detection on Fedora 22 (Ira Cooper)
- build: fix pg ref disabling (William A. Kennington III)
- build: fix ppc build (James Page)

- build: install-deps: misc fixes (Loic Dachary)
- build: install-deps.sh improvements (Loic Dachary)
- build: install-deps: support OpenSUSE (Loic Dachary)
- build: make\_dist\_tarball.sh (Sage Weil)
- build: many cmake improvements
- build: misc cmake fixes (Matt Benjamin)
- build: misc fixes (Boris Ranto, Ken Dreyer, Owen Synge)
- build: OSX build fixes (Yan, Zheng)
- build: remove rest-bench
- ceph-authtool: fix return code on error (Gerhard Muntingh)
- ceph-detect-init: added Linux Mint (Michal Jarzabek)
- ceph-detect-init: robust init system detection (Owen Synge)
- ceph-disk: ensure ‘zap’ only operates on a full disk (#11272 Loic Dachary)
- ceph-disk: fix zap sgdisk invocation (Owen Synge, Thorsten Behrens)
- ceph-disk: follow ceph-osd hints when creating journal (#9580 Sage Weil)
- ceph-disk: handle re-using existing partition (#10987 Loic Dachary)
- ceph-disk: improve parted output parsing (#10983 Loic Dachary)
- ceph-disk: install pip > 6.1 (#11952 Loic Dachary)
- ceph-disk: make suppression work for activate-all and activate-journal (Dan van der Ster)
- ceph-disk: many fixes (Loic Dachary, Alfredo Deza)
- ceph-disk: fixes to respect init system (Loic Dachary, Owen Synge)
- ceph-disk: pass -cluster arg on prepare subcommand (Kefu Chai)
- ceph-disk: support for multipath devices (Loic Dachary)
- ceph-disk: support NVMe device partitions (#11612 Ilja Slepnev)
- ceph: fix ‘df’ units (Zhe Zhang)
- ceph: fix parsing in interactive cli mode (#11279 Kefu Chai)
- cephfs-data-scan: many additions, improvements (John Spray)

- ceph-fuse: do not require successful remount when unmounting (#10982 Greg Farnum)
- ceph-fuse, libcephfs: don't clear COMPLETE when trimming null (Yan, Zheng)
- ceph-fuse, libcephfs: drop inode when rmdir finishes (#11339 Yan, Zheng)
- ceph-fuse, libcephfs: fix uninlne (#11356 Yan, Zheng)
- ceph-fuse, libcephfs: hold exclusive caps on dirs we "own" (#11226 Greg Farnum)
- ceph-fuse: mostly behave on 32-bit hosts (Yan, Zheng)
- ceph: improve error output for 'tell' (#11101 Kefu Chai)
- ceph-monstore-tool: fix store-copy (Huangjun)
- ceph: new 'ceph daemonperf' command (John Spray, Mykola Golub)
- ceph-objectstore-tool: many many improvements (David Zafman)
- ceph-objectstore-tool: refactoring and cleanup (John Spray)
- ceph-post-file: misc fixes (Joey McDonald, Sage Weil)
- ceph\_test\_rados: test pipelined reads (Zhiqiang Wang)
- client: avoid sending unnecessary FLUSHSNAP messages (Yan, Zheng)
- client: exclude setfilelock when calculating oldest tid (Yan, Zheng)
- client: fix error handling in check\_pool\_perm (John Spray)
- client: fsync waits only for inode's caps to flush (Yan, Zheng)
- client: invalidate kernel dcache when cache size exceeds limits (Yan, Zheng)
- client: make fsync wait for unsafe dir operations (Yan, Zheng)
- client: pin lookup dentry to avoid inode being freed (Yan, Zheng)
- common: add descriptions to perfcounters (Kiseleva Alyona)
- common: add perf counter descriptions (Alyona Kiseleva)
- common: bufferlist performance tuning (Piotr Dalek, Sage Weil)
- common: detect overflow of int config values (#11484 Kefu Chai)
- common: fix bit\_vector extent calc (#12611 Jason Dillaman)
- common: fix json parsing of utf8 (#7387 Tim Serong)
- common: fix leak of pthread\_mutexattr (#11762 Ketor Meng)

- common: fix LTTNG vs fork issue (Josh Durgin)
- common: fix throttle max change (Henry Chang)
- common: make mutex more efficient
- common: make work queue addition/removal thread safe (#12662 Jason Dillaman)
- common: optracker improvements (Zhiqiang Wang, Jianpeng Ma)
- common: PriorityQueue tests (Kefu Chai)
- common: some async compression infrastructure (Haomai Wang)
- crush: add -check to validate dangling names, max osd id (Kefu Chai)
- crush: cleanup, sync with kernel (Ilya Dryomov)
- crush: fix crash from invalid ‘take’ argument (#11602 Shiva Rkreddy, Sage Weil)
- crush: fix divide-by-2 in straw2 (#11357 Yann Dupont, Sage Weil)
- crush: fix has\_v4\_buckets (#11364 Sage Weil)
- crush: fix subtree base weight on adjust\_subtree\_weight (#11855 Sage Weil)
- crush: respect default replicated ruleset config on map creation (Ilya Dryomov)
- crushtool: fix order of operations, usage (Sage Weil)
- crypto: fix NSS leak (Jason Dillaman)
- crypto: fix unbalanced init/shutdown (#12598 Zheng Yan)
- deb: fix rest-bench-dbg and ceph-test-dbg dependendies (Ken Dreyer)
- debian: minor package reorg (Ken Dreyer)
- deb, rpm: move ceph-objectstore-tool to ceph (Ken Dreyer)
- doc: docuemnt object corpus generation (#11099 Alexis Normand)
- doc: document region hostnames (Robin H. Johnson)
- doc: fix gender neutrality (Alexandre Maragone)
- doc: fix install doc (#10957 Kefu Chai)
- doc: fix sphinx issues (Kefu Chai)
- doc: man page updates (Kefu Chai)
- doc: mds data structure docs (Yan, Zheng)

- doc: misc updates (Francois Lafont, Ken Dreyer, Kefu Chai, Owen Synge, Gael Fenet-Garde, Loic Dachary, Yannick Atchy-Dalama, Jiaying Ren, Kevin Caradant, Robert Maxime, Nicolas Yong, Germain Chipaux, Arthur Gorjux, Gabriel Sentucq, Clement Lebrun, Jean-Remi Deveaux, Clair Massot, Robin Tang, Thomas Laumondais, Jordan Dorne, Yuan Zhou, Valentin Thomas, Pierre Chaumont, Benjamin Troquereau, Benjamin Sesia, Vikhyat Umrao, Nilamdyuti Goswami, Vartika Rai, Florian Haas, Loic Dachary, Simon Guinot, Andy Allan, Alistair Israel, Ken Dreyer, Robin Rehu, Lee Revell, Florian Marsylle, Thomas Johnson, Bosse Klykken, Travis Rhoden, Ian Kelling)
- doc: swift tempurls (#10184 Abhishek Lekshmanan)
- doc: switch doxygen integration back to breathe (#6115 Kefu Chai)
- doc: update release schedule docs (Loic Dachary)
- erasure-code: cleanup (Kefu Chai)
- erasure-code: improve tests (Loic Dachary)
- erasure-code: shec: fix recovery bugs (Takanori Nakao, Shotaro Kawaguchi)
- erasure-code: update ISA-L to 2.13 (Yuan Zhou)
- gmock: switch to submodule (Danny Al-Gaaf, Loic Dachary)
- hadoop: add terasort test (Noah Watkins)
- init-radosgw: merge with sysv version; fix enumeration (Sage Weil)
- java: fix libcephfs bindings (Noah Watkins)
- libcephfs: add pread, pwrite (Jevon Qiao)
- libcephfs,ceph-fuse: cache cleanup (Zheng Yan)
- libcephfs,ceph-fuse: fix request resend on cap reconnect (#10912 Yan, Zheng)
- librados: add config observer (Alistair Strachan)
- librados: add FULL\_TRY and FULL\_FORCE flags for dealing with full clusters or pools (Sage Weil)
- librados: add src\_fadvise\_flags for copy-from (Jianpeng Ma)
- librados: define C++ flags from C constants (Josh Durgin)
- librados: fadvise flags per op (Jianpeng Ma)
- librados: fix last\_force\_resent handling (#11026 Jianpeng Ma)
- librados: fix memory leak from C\_TwoContexts (Xiong Yiliang)

- librados: fix notify completion race (#13114 Sage Weil)
- librados: fix striper when stripe\_count = 1 and stripe\_unit != object\_size (#11120 Yan, Zheng)
- librados, libcephfs: randomize client nonces (Josh Durgin)
- librados: op perf counters (John Spray)
- librados: pybind: fix binary omap values (Robin H. Johnson)
- librados: pybind: fix write() method return code (Javier Guerra)
- librados: respect default\_crush\_ruleset on pool\_create (#11640 Yuan Zhou)
- libradosstriper: fix leak (Danny Al-Gaaf)
- librbd: add const for single-client-only features (Josh Durgin)
- librbd: add deep-flatten operation (Jason Dillaman)
- librbd: add purge\_on\_error cache behavior (Jianpeng Ma)
- librbd: allow additional metadata to be stored with the image (Haomai Wang)
- librbd: avoid blocking aio API methods (#11056 Jason Dillaman)
- librbd: better handling for dup flatten requests (#11370 Jason Dillaman)
- librbd: cancel in-flight ops on watch error (#11363 Jason Dillaman)
- librbd: default new images to format 2 (#11348 Jason Dillaman)
- librbd: fadvise for copy, export, import (Jianpeng Ma)
- librbd: fast diff implementation that leverages object map (Jason Dillaman)
- librbd: fix fast diff bugs (#11553 Jason Dillaman)
- librbd: fix image format detection (Zhiqiang Wang)
- librbd: fix lock ordering issue (#11577 Jason Dillaman)
- librbd: fix reads larger than the cache size (Lu Shi)
- librbd: fix snapshot creation when other snap is active (#11475 Jason Dillaman)
- librbd: flatten/copyup fixes (Jason Dillaman)
- librbd: handle NOCACHE fadvise flag (Jinapeng Ma)
- librbd: lockdep, helgrind validation (Jason Dillaman, Josh Durgin)
- librbd: metadata filter fixes (Haomai Wang)

- librbd: misc aio fixes (#5488 Jason Dillaman)
- librbd: misc rbd fixes (#11478 #11113 #11342 #11380 Jason Dillaman, Zhiqiang Wang)
- librbd: new diff\_iterate2 API (Jason Dillaman)
- librbd: object map rebuild support (Jason Dillaman)
- librbd: only update image flags while hold exclusive lock (#11791 Jason Dillaman)
- librbd: optionally disable allocation hint (Haomai Wang)
- librbd: prevent race between resize requests (#12664 Jason Dillaman)
- librbd: readahead fixes (Zhiqiang Wang)
- librbd: return result code from close (#12069 Jason Dillaman)
- librbd: store metadata, including config options, in image (Haomai Wang)
- librbd: tolerate old osds when getting image metadata (#11549 Jason Dillaman)
- librbd: use write\_full when possible (Zhiqiang Wang)
- log: fix data corruption race resulting from log rotation (#12465 Samuel Just)
- logrotate.d: prefer service over invoke-rc.d (#11330 Win Hierman, Sage Weil)
- mds: add 'damaged' state to MDSMap (John Spray)
- mds: add nicknames for perfcounters (John Spray)
- mds: avoid emitting cap warnigns before evicting session (John Spray)
- mds: avoid getting stuck in XLOCKDONE (#11254 Yan, Zheng)
- mds: disable problematic rstat propagation into snap parents (Yan, Zheng)
- mds: do not add snapped items to bloom filter (Yan, Zheng)
- mds: expose frags via asok (John Spray)
- mds: fix expected holes in journal objects (#13167 Yan, Zheng)
- mds: fix handling for missing mydir dirfrag (#11641 John Spray)
- mds: fix integer truncation on large client ids (Henry Chang)
- mds: fix mydir replica issue with shutdown (#10743 John Spray)
- mds: fix out-of-order messages (#11258 Yan, Zheng)
- mds: fix rejoin (Yan, Zheng)

- mds: fix setting entire file layout in one setxattr (John Spray)
- mds: fix shutdown (John Spray)
- mds: fix shutdown with strays (#10744 John Spray)
- mds: fix SnapServer crash on deleted pool (John Spray)
- mds: fix snapshot bugs (Yan, Zheng)
- mds: fix stray reintegration (Yan, Zheng)
- mds: fix stray handling (John Spray)
- mds: fix suicide beacon (John Spray)
- mds: flush immediately in do\_open\_truncate (#11011 John Spray)
- mds: handle misc corruption issues (John Spray)
- mds: improve dump methods (John Spray)
- mds: many fixes (Yan, Zheng, John Spray, Greg Farnum)
- mds: many snapshot and stray fixes (Yan, Zheng)
- mds: misc fixes (Jianpeng Ma, Dan van der Ster, Zhang Zhi)
- mds: misc journal cleanups and fixes (#10368 John Spray)
- mds: misc repair improvements (John Spray)
- mds: misc snap fixes (Zheng Yan)
- mds: misc snapshot fixes (Yan, Zheng)
- mds: new SessionMap storage using omap (#10649 John Spray)
- mds: persist completed\_requests reliably (#11048 John Spray)
- mds: reduce memory consumption (Yan, Zheng)
- mds: respawn instead of suicide on blacklist (John Spray)
- mds: separate safe\_pos in Journaler (#10368 John Spray)
- mds: snapshot rename support (#3645 Yan, Zheng)
- mds: store layout on header object (#4161 John Spray)
- mds: throttle purge stray operations (#10390 John Spray)
- mds: tolerate clock jumping backwards (#11053 Yan, Zheng)

- mds: warn when clients fail to advance oldest\_client\_tid (#10657 Yan, Zheng)
- misc cleanups and fixes (Danny Al-Gaaf)
- misc coverity fixes (Danny Al-Gaaf)
- misc performance and cleanup (Nathan Cutler, Xinxin Shu)
- mon: add cache over MonitorDBStore (Kefu Chai)
- mon: add 'mon\_metadata <id>' command (Kefu Chai)
- mon: add 'node ls ...' command (Kefu Chai)
- mon: add NOFORWARD, OBSOLETE, DEPRECATE flags for mon commands (Joao Eduardo Luis)
- mon: add PG count to 'ceph osd df' output (Michal Jarzabek)
- mon: 'ceph osd metadata' can dump all osds (Haomai Wang)
- mon: clean up, reorg some mon commands (Joao Eduardo Luis)
- monclient: flush\_log (John Spray)
- mon: detect kv backend failures (Sage Weil)
- mon: disallow >2 tiers (#11840 Kefu Chai)
- mon: disallow ec pools as tiers (#11650 Samuel Just)
- mon: do not deactivate last mds (#10862 John Spray)
- mon: fix average utilization calc for 'osd df' (Mykola Golub)
- mon: fix CRUSH map test for new pools (Sage Weil)
- mon: fix log dump crash when debugging (Mykola Golub)
- mon: fix mds beacon replies (#11590 Kefu Chai)
- mon: fix metadata update race (Mykola Golub)
- mon: fix min\_last\_epoch\_clean tracking (Kefu Chai)
- mon: fix 'pg ls' sort order, state names (#11569 Kefu Chai)
- mon: fix refresh (#11470 Joao Eduardo Luis)
- mon: fix variance calc in 'osd df' (Sage Weil)
- mon: improve callout to crushtool (Mykola Golub)
- mon: make blocked op messages more readable (Jianpeng Ma)

- mon: make osd get pool 'all' only return applicable fields (#10891 Michal Jarzabek)
- mon: misc scaling fixes (Sage Weil)
- mon: normalize erasure-code profile for storage and comparison (Loic Dachary)
- mon: only send mon metadata to supporting peers (Sage Weil)
- mon: optionally specify osd id on 'osd create' (Mykola Golub)
- mon: 'osd tree' fixes (Kefu Chai)
- mon: periodic background scrub (Joao Eduardo Luis)
- mon: prevent bucket deletion when referenced by a crush rule (#11602 Sage Weil)
- mon: prevent pgp\_num > pg\_num (#12025 Xinxin Shu)
- mon: prevent pool with snapshot state from being used as a tier (#11493 Sage Weil)
- mon: prime pg\_temp when CRUSH map changes (Sage Weil)
- mon: refine check\_remove\_tier checks (#11504 John Spray)
- mon: reject large max\_mds values (#12222 John Spray)
- mon: remove spurious who arg from 'mds rm ...' (John Spray)
- mon: streamline session handling, fix memory leaks (Sage Weil)
- mon: upgrades must pass through hammer (Sage Weil)
- mon: warn on bogus cache tier config (Jianpeng Ma)
- msgr: add ceph\_perf\_msgr tool (Hoamai Wang)
- msgr: async: fix seq handling (Haomai Wang)
- msgr: async: many many fixes (Haomai Wang)
- msgr: simple: fix clear\_pipe (#11381 Haomai Wang)
- msgr: simple: fix connect\_seq assert (Haomai Wang)
- msgr: xio: fastpath improvements (Raju Kurunkad)
- msgr: xio: fix ip and nonce (Raju Kurunkad)
- msgr: xio: improve lane assignment (Vu Pham)
- msgr: xio: sync with accellio v1.4 (Vu Pham)

- msgr: xio: misc fixes (#10735 Matt Benjamin, Kefu Chai, Danny Al-Gaaf, Raju Kurunkad, Vu Pham, Casey Bodley)
- msg: unit tests (Haomai Wang)
- objectcacher: misc bug fixes (Jianpeng Ma)
- osd: add latency perf counters for tier operations (Xinze Chi)
- osd: add misc perfcounters (Xinze Chi)
- osd: add simple sleep injection in recovery (Sage Weil)
- osd: allow SEEK\_HOLE/SEEK\_DATA for sparse read (Zhiqiang Wang)
- osd: avoid dup omap sets for in pg metadata (Sage Weil)
- osd: avoid multiple hit set insertions (Zhiqiang Wang)
- osd: avoid transaction append in some cases (Sage Weil)
- osd: break PG removal into multiple iterations (#10198 Guang Yang)
- osd: cache proxy-write support (Zhiqiang Wang, Samuel Just)
- osd: check scrub state when handling map (Jianpeng Ma)
- osd: clean up some constness, privateness (Kefu Chai)
- osd: clean up temp object if promotion fails (Jianpeng Ma)
- osd: configure promotion based on write recency (Zhiqiang Wang)
- osd: constrain collections to meta and PGs (normal and temp) (Sage Weil)
- osd: don't send dup MMonGetOSDMap requests (Sage Weil, Kefu Chai)
- osd: EIO injection (David Zhang)
- osd: elminiate txn apend, ECSubWrite copy (Samuel Just)
- osd: erasure-code: drop entries according to LRU (Andreas-Joachim Peters)
- osd: erasure-code: fix SHEC floating point bug (#12936 Loic Dachary)
- osd: erasure-code: update to ISA-L 2.14 (Yuan Zhou)
- osd: filejournal: cleanup (David Zafman)
- osd: filestore: clone using splice (Jianpeng Ma)
- osd: filestore: fix recursive lock (Xinxin Shu)
- osd: fix check\_for\_full (Henry Chang)

- osd: fix dirty accounting in make\_writeable (Zhiqiang Wang)
- osd: fix dup promotion lost op bug (Zhiqiang Wang)
- osd: fix endless repair when object is unrecoverable (Jianpeng Ma, Kefu Chai)
- osd: fix hitset object naming to use GMT (Kefu Chai)
- osd: fix misc memory leaks (Sage Weil)
- osd: fix negative degraded stats during backfill (Guang Yang)
- osd: fix osdmap dump of blacklist items (John Spray)
- osd: fix peek\_queue locking in FileStore (Xinze Chi)
- osd: fix pg resurrection (#11429 Samuel Just)
- osd: fix promotion vs full cache tier (Samuel Just)
- osd: fix replay requeue when pg is still activating (#13116 Samuel Just)
- osd: fix scrub stat bugs (Sage Weil, Samuel Just)
- osd: fix snap flushing from cache tier (again) (#11787 Samuel Just)
- osd: fix snap handling on promotion (#11296 Sam Just)
- osd: fix temp-clearing (David Zafman)
- osd: force promotion for ops EC can't handle (Zhiqiang Wang)
- osd: handle log split with overlapping entries (#11358 Samuel Just)
- osd: ignore non-existent osds in unfound calc (#10976 Mykola Golub)
- osd: improve behavior on machines with large memory pages (Steve Capper)
- osd: include a temp namespace within each collection/pgid (Sage Weil)
- osd: increase default max open files (Owen Synge)
- osd: keyvaluestore: misc fixes (Varada Kari)
- osd: low and high speed flush modes (Mingxin Liu)
- osd: make suicide timeouts individually configurable (Samuel Just)
- osd: merge multiple setattr calls into a setattrs call (Xinxin Shu)
- osd: misc fixes (Ning Yao, Kefu Chai, Xinze Chi, Zhiqiang Wang, Jianpeng Ma)
- osd: move scrub in OpWQ (Samuel Just)

- osd: newstore prototype (Sage Weil)
- osd: ObjectStore internal API refactor (Sage Weil)
- osd: peer\_features includes self (David Zafman)
- osd: pool size change triggers new interval (#11771 Samuel Just)
- osd: prepopulate needs\_recovery\_map when only one peer has missing (#9558 Guang Yang)
- osd: randomize scrub times (#10973 Kefu Chai)
- osd: recovery, peering fixes (#11687 Samuel Just)
- osd: refactor scrub and digest recording (Sage Weil)
- osd: refuse first write to EC object at non-zero offset (Jianpeng Ma)
- osd: relax reply order on proxy read (#11211 Zhiqiang Wang)
- osd: require firefly features (David Zafman)
- osd: set initial crush weight with more precision (Sage Weil)
- osd: SHEC no longer experimental
- osd: skip promotion for flush/evict op (Zhiqiang Wang)
- osd: stripe over small xattrs to fit in XFS's 255 byte inline limit (Sage Weil, Ning Yao)
- osd: sync object\_map on syncfs (Samuel Just)
- osd: take excl lock of op is rw (Samuel Just)
- osd: throttle evict ops (Yunchuan Wen)
- osd: upgrades must pass through hammer (Sage Weil)
- osd: use a temp object for recovery (Sage Weil)
- osd: use blkid to collection partition information (Joseph Handzik)
- osd: use SEEK\_HOLE / SEEK\_DATA for sparse copy (Xinxin Shu)
- osd: WBThrottle cleanups (Jianpeng Ma)
- osd: write journal header on clean shutdown (Xinze Chi)
- osdc/Objecter: allow per-pool calls to op\_cancel\_writes (John Spray)
- os/filestore: enlarge getxattr buffer size (Jianpeng Ma)

- pybind: pep8 cleanups (Danny Al-Gaaf)
- pycephfs: many fixes for bindings (Haomai Wang)
- qa: fix filelock\_interrupt.py test (Yan, Zheng)
- qa: improve ceph-disk tests (Loic Dachary)
- qa: improve docker build layers (Loic Dachary)
- qa: run-make-check.sh script (Loic Dachary)
- rados: add -striper option to use libradosstriper (#10759 Sebastien Ponce)
- rados: bench: add -no-verify option to improve performance (Piotr Dalek)
- rados bench: misc fixes (Dmitry Yatsushkevich)
- rados: fix error message on failed pool removal (Wido den Hollander)
- radosgw-admin: add ‘bucket check’ function to repair bucket index (Yehuda Sadeh)
- radosgw-admin: fix subuser modify output (#12286 Guce)
- rados: handle -snapid arg properly (Abhishek Lekshmanan)
- rados: improve bench buffer handling, performance (Piotr Dalek)
- rados: misc bench fixes (Dmitry Yatsushkevich)
- rados: new pool import implementation (John Spray)
- rados: translate errno to string in CLI (#10877 Kefu Chai)
- rbd: accept map options config option (Ilya Dryomov)
- rbd: add disk usage tool (#7746 Jason Dillaman)
- rbd: allow unmapping by spec (Ilya Dryomov)
- rbd: cli: fix arg parsing with -io-pattern (Dmitry Yatsushkevich)
- rbd: deprecate -new-format option (Jason Dillman)
- rbd: fix error messages (#2862 Rajesh Nambiar)
- rbd: fix link issues (Jason Dillaman)
- rbd: improve CLI arg parsing, usage (Ilya Dryomov)
- rbd: rbd-replay-prep and rbd-replay improvements (Jason Dillaman)
- rbd: recognize queue\_depth kernel option (Ilya Dryomov)

- rbd: support G and T units for CLI (Abhishek Lekshmanan)
- rbd: update rbd man page (Ilya Dryomov)
- rbd: update xfstests tests (Douglas Fuller)
- rbd: use image-spec and snap-spec in help (Vikhyat Umrao, Ilya Dryomov)
- rest-bench: misc fixes (Shawn Chen)
- rest-bench: support https (#3968 Yuan Zhou)
- rgw: add max multipart upload parts (#12146 Abhishek Dixit)
- rgw: add missing headers to Swift container details (#10666 Ahmad Faheem, Dmytro Iurchenko)
- rgw: add stats to headers for account GET (#10684 Yuan Zhou)
- rgw: add Transaction-Id to response (Abhishek Dixit)
- rgw: add X-Timestamp for Swift containers (#10938 Radoslaw Zarzynski)
- rgw: always check if token is expired (#11367 Anton Aksola, Riku Lehto)
- rgw: conversion tool to repair broken multipart objects (#12079 Yehuda Sadeh)
- rgw: document layout of pools and objects (Pete Zaitcev)
- rgw: do not enclose bucket header in quotes (#11860 Wido den Hollander)
- rgw: do not prefetch data for HEAD requests (Guang Yang)
- rgw: do not preserve ACLs when copying object (#12370 Yehuda Sadeh)
- rgw: do not set content-type if length is 0 (#11091 Orit Wasserman)
- rgw: don't clobber bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: don't use end\_marker for namespaced object listing (#11437 Yehuda Sadeh)
- rgw: don't use rgw\_socket\_path if frontend is configured (#11160 Yehuda Sadeh)
- rgw: enforce Content-Length for POST on Swift cont/obj (#10661 Radoslaw Zarzynski)
- rgw: error out if frontend did not send all data (#11851 Yehuda Sadeh)
- rgw: expose the number of unhealthy workers through admin socket (Guang Yang)
- rgw: fail if parts not specified on multipart upload (#11435 Yehuda Sadeh)
- rgw: fix assignment of copy obj attributes (#11563 Yehuda Sadeh)

- rgw: fix broken stats in container listing (#11285 Radoslaw Zarzynski)
- rgw: fix bug in domain/subdomain splitting (Robin H. Johnson)
- rgw: fix casing of Content-Type header (Robin H. Johnson)
- rgw: fix civetweb max threads (#10243 Yehuda Sadeh)
- rgw: fix Connection: header handling (#12298 Wido den Hollander)
- rgw: fix copy metadata, support X-Copied-From for swift (#10663 Radoslaw Zarzynski)
- rgw: fix data corruptions race condition (#11749 Wuxingyi)
- rgw: fix decoding of X-Object-Manifest from GET on Swift DLO (Radoslaw Rzarzynski)
- rgw: fix GET on swift account when limit == 0 (#10683 Radoslaw Zarzynski)
- rgw: fix handling empty metadata items on Swift container (#11088 Radoslaw Zarzynski)
- rgw: fix JSON response when getting user quota (#12117 Wuxingyi)
- rgw: fix locator for objects starting with \_ (#11442 Yehuda Sadeh)
- rgw: fix log rotation (Wuxingyi)
- rgw: fix multipart upload in retry path (#11604 Yehuda Sadeh)
- rgw: fix quota enforcement on POST (#11323 Sergey Arkhipov)
- rgw: fix reset\_loc (#11974 Yehuda Sadeh)
- rgw: fix return code on missing upload (#11436 Yehuda Sadeh)
- rgw: fix sysvinit script
- rgw: fix sysvinit script w/ multiple instances (Sage Weil, Pavan Rallabhandi)
- rgw: force content\_type for swift bucket stats requests (#12095 Orit Wasserman)
- rgw: force content type header on responses with no body (#11438 Orit Wasserman)
- rgw: generate Date header for civetweb (#10873 Radoslaw Zarzynski)
- rgw: generate new object tag when setting attrs (#11256 Yehuda Sadeh)
- rgw: improve content-length env var handling (#11419 Robin H. Johnson)
- rgw: improved support for swift account metadata (Radoslaw Zarzynski)
- rgw: improve handling of already removed buckets in expirer (Radoslaw Rzarzynski)

- rgw: issue aio for first chunk before flush cached data (#11322 Guang Yang)
- rgw: log to /var/log/ceph instead of /var/log/radosgw
- rgw: make init script wait for radosgw to stop (#11140 Dmitry Yatsushkevich)
- rgw: make max put size configurable (#6999 Yuan Zhou)
- rgw: make quota/gc threads configurable (#11047 Guang Yang)
- rgw: make read user buckets backward compat (#10683 Radoslaw Zarzynski)
- rgw: merge manifests properly with prefix override (#11622 Yehuda Sadeh)
- rgw: only scan for objects not in a namespace (#11984 Yehuda Sadeh)
- rgw: orphan detection tool (Yehuda Sadeh)
- rgw: pass in civetweb configurables (#10907 Yehuda Sadeh)
- rgw: rectify 202 Accepted in PUT response (#11148 Radoslaw Zarzynski)
- rgw: remove meta file after deleting bucket (#11149 Orit Wasserman)
- rgw: remove trailing :port from HTTP\_HOST header (Sage Weil)
- rgw: return 412 on bad limit when listing buckets (#11613 Yehuda Sadeh)
- rgw: rework X-Trans-Id header to conform with Swift API (Radoslaw Rzarzynski)
- rgw: s3 encoding-type for get bucket (Jeff Weber)
- rgw: send ETag, Last-Modified for swift (#11087 Radoslaw Zarzynski)
- rgw: set content length on container GET, PUT, DELETE, HEAD (#10971, #11036 Radoslaw Zarzynski)
- rgw: set max buckets per user in ceph.conf (Vikhyat Umrao)
- rgw: shard work over multiple librados instances (Pavan Rallabhandi)
- rgw: support end marker on swift container GET (#10682 Radoslaw Zarzynski)
- rgw: support for Swift expiration API (Radoslaw Rzarzynski, Yehuda Sadeh)
- rgw: swift: allow setting attributes with COPY (#10662 Ahmad Faheem, Dmytro Iurchenko)
- rgw: swift: do not override sent content type (#12363 Orit Wasserman)
- rgw: swift: enforce Content-Type in response (#12157 Radoslaw Zarzynski)
- rgw: swift: fix account listing (#11501 Radoslaw Zarzynski)

- rgw: swift: fix metadata handling on copy (#10645 Radoslaw Zarzynski)
- rgw: swift: send Last-Modified header (#10650 Radoslaw Zarzynski)
- rgw: swift: set Content-Length for account GET (#12158 Radoslav Zarzynski)
- rgw: swift: set content-length on keystone tokens (#11473 Herv Rousseau)
- rgw: update keystone cache with token info (#11125 Yehuda Sadeh)
- rgw: update to latest civetweb, enable config for IPv6 (#10965 Yehuda Sadeh)
- rgw: use attrs from source bucket on copy (#11639 Javier M. Mellid)
- rgw: use correct oid for gc chains (#11447 Yehuda Sadeh)
- rgw: user rm is idempotent (Orit Wasserman)
- rgw: use unique request id for civetweb (#10295 Orit Wasserman)
- rocksdb: add perf counters for get/put latency (Xinxin Shu)
- rocksdb, leveldb: fix compact\_on\_mount (Xiaoxi Chen)
- rocksdb: pass options as single string (Xiaoxi Chen)
- rocksdb: update to latest (Xiaoxi Chen)
- rpm: add suse firewall files (Tim Serong)
- rpm: always rebuild and install man pages for rpm (Owen Synge)
- rpm: loosen ceph-test dependencies (Ken Dreyer)
- rpm: many spec file fixes (Owen Synge, Ken Dreyer)
- rpm: misc fixes (Boris Ranto, Owen Synge, Ken Dreyer, Ira Cooper)
- rpm: misc systemd and SUSE fixes (Owen Synge, Nathan Cutler)
- selinux policy (Boris Ranto, Milan Broz)
- systemd: logrotate fixes (Tim Serong, Lars Marowsky-Bree, Nathan Cutler)
- systemd: many fixes (Sage Weil, Owen Synge, Boris Ranto, Dan van der Ster)
- systemd: run daemons as user ceph
- sysvinit compat: misc fixes (Owen Synge)
- test: misc fs test improvements (John Spray, Loic Dachary)
- test: python tests, linter cleanup (Alfredo Deza)

- tests: fixes for rbd xtests (Douglas Fuller)
- tests: fix tiering health checks (Loic Dachary)
- tests for low-level performance (Haomai Wang)
- tests: many ec non-regression improvements (Loic Dachary)
- tests: many many ec test improvements (Loic Dachary)
- upstart: throttle restarts (#11798 Sage Weil, Greg Farnum)

## v9.1.0 Infernalis release candidate

This is the first Infernalis release candidate. There have been some major changes since Hammer, and the upgrade process is non-trivial. Please read carefully.

### Getting the release candidate

The v9.1.0 packages are pushed to the development release repositories:

1. <http://download.ceph.com/rpm-testing>
2. <http://download.ceph.com/debian-testing>

For for info, see:

1. <http://docs.ceph.com/docs/master/install/get-packages/>

Or install with ceph-deploy via:

1. `ceph-deploy install --testing HOST`

### Known issues

- librbd and librados ABI compatibility is broken. Be careful installing this RC on client machines (e.g., those running qemu). It will be fixed in the final v9.2.0 release.

### Major Changes from Hammer

- *General:*
  - Ceph daemons are now managed via systemd (with the exception of Ubuntu Trusty, which still uses upstart).

- Ceph daemons run as ‘ceph’ user instead of root.
- On Red Hat distros, there is also an SELinux policy.
- *RADOS*:
  - The RADOS cache tier can now proxy write operations to the base tier, allowing writes to be handled without forcing migration of an object into the cache.
  - The SHEC erasure coding support is no longer flagged as experimental. SHEC trades some additional storage space for faster repair.
  - There is now a unified queue (and thus prioritization) of client IO, scrubbing, and snapshot trimming.
  - There have been many improvements to low-level repair tooling (`ceph-objectstore-tool`).
  - The internal ObjectStore API has been significantly cleaned up in order to facilitate new storage backends like NewStore.
- *RGW*:
  - The Swift API now supports object expiration.
  - There are many Swift API compatibility improvements.
- *RBD*:
  - The `rbd du` command shows actual usage (quickly, when object-map is enabled).
  - The object-map feature has seen many stability improvements.
  - Object-map and exclusive-lock features can be enabled or disabled dynamically.
  - You can now store user metadata and set persistent librbd options associated with individual images.
  - The new deep-flatten features allows flattening of a clone and all of its snapshots. (Previously snapshots could not be flattened.)
  - The export-diff command command is now faster (it uses aio). There is also a new fast-diff feature.
  - The `-size` argument can be specified with a suffix for units (e.g., `--size 64G`).
  - There is a new `rbd status` command that, for now, shows who has the image open/mapped.

- *CephFS*:
  - You can now rename snapshots.
  - There have been ongoing improvements around administration, diagnostics, and the check and repair tools.
  - The caching and revocation of client cache state due to unused inodes has been dramatically improved.
  - The ceph-fuse client behaves better on 32-bit hosts.

## Distro compatibility

---

We have decided to drop support for many older distributions so that we can move to a newer compiler toolchain (e.g., C++11). Although it is still possible to build Ceph on older distributions by installing backported development tools, we are not building and publishing release packages for them on ceph.com.

In particular,

- CentOS 7 or later; we have dropped support for CentOS 6 (and other RHEL 6 derivatives, like Scientific Linux 6).
- Debian Jessie 8.x or later; Debian Wheezy 7.x's g++ has incomplete support for C++11 (and no systemd).
- Ubuntu Trusty 14.04 or later; Ubuntu Precise 12.04 is no longer supported.
- Fedora 22 or later.

## Upgrading from Firefly

---

Upgrading directly from Firefly v0.80.z is not possible. All clusters must first upgrade to Hammer v0.94.4 or a later v0.94.z release; only then is it possible to do online upgrade to Infernalis 9.2.z.

User can upgrade to latest hammer v0.94.z from gitbuilder with(also refer the hammer release notes for more details):

```
1. ceph-deploy install --release hammer HOST
```

## Upgrading from Hammer

---

- All cluster nodes must first upgrade to Hammer v0.94.4 or a later v0.94.z release; only then is it possible to do online upgrade to Infernalis 9.2.z.

- For all distributions that support systemd (CentOS 7, Fedora, Debian Jessie 8.x, OpenSUSE), ceph daemons are now managed using native systemd files instead of the legacy sysvinit scripts. For example:

```
1. systemctl start ceph.target      # start all daemons
2. systemctl status ceph-osd@12     # check status of osd.12
```

The main notable distro that is *not* yet using systemd is Ubuntu trusty 14.04.  
(The next Ubuntu LTS, 16.04, will use systemd instead of upstart.)

- Ceph daemons now run as user and group `ceph` by default. The `ceph` user has a static UID assigned by Fedora and Debian (also used by derivative distributions like RHEL/CentOS and Ubuntu). On SUSE the `ceph` user will currently get a dynamically assigned UID when the user is created.

If your systems already have a `ceph` user, the package upgrade process will usually fail with an error. We suggest you first remove or rename the existing '`ceph`' user and then upgrade.

When upgrading, administrators have two options:

i. Add the following line to `ceph.conf` on all hosts:

```
1. setuser match path = /var/lib/ceph/$type/$cluster-$id
```

This will make the Ceph daemons run as root (i.e., not drop privileges and switch to user `ceph`) if the daemon's data directory is still owned by root. Newly deployed daemons will be created with data owned by user `ceph` and will run with reduced privileges, but upgraded daemons will continue to run as root.

ii. Fix the data ownership during the upgrade. This is the preferred option, but is more work. The process for each host would be to:

a. Upgrade the `ceph` package. This creates the `ceph` user and group. For example:

```
1. ceph-deploy install --stable infernalis HOST
```

b. Stop the daemon(s):

```
1. service ceph stop          # fedora, centos, rhel, debian
2. stop ceph-all              # ubuntu
```

c. Fix the ownership:

```
1. chown -R ceph:ceph /var/lib/ceph
2. chown -R ceph:ceph /var/log/ceph
```

d. Restart the daemon(s):

```

1. start ceph-all          # ubuntu
2. systemctl start ceph.target # debian, centos, fedora, rhel

```

- The on-disk format for the experimental KeyValueStore OSD backend has changed. You will need to remove any OSDs using that backend before you upgrade any test clusters that use it.

## Upgrade notes

- When a pool quota is reached, librados operations now block indefinitely, the same way they do when the cluster fills up. (Previously they would return -ENOSPC). By default, a full cluster or pool will now block. If your librados application can handle ENOSPC or EDQUOT errors gracefully, you can get error returns instead by using the new librados OPERATION\_FULL\_TRY flag.

## Notable changes

NOTE: These notes are somewhat abbreviated while we find a less time-consuming process for generating them.

- build: C++11 now supported
- build: many cmake improvements
- build: OSX build fixes (Yan, Zheng)
- build: remove rest-bench
- ceph-disk: many fixes (Loic Dachary)
- ceph-disk: support for multipath devices (Loic Dachary)
- ceph-fuse: mostly behave on 32-bit hosts (Yan, Zheng)
- ceph-objectstore-tool: many improvements (David Zafman)
- common: bufferlist performance tuning (Piotr Dalek, Sage Weil)
- common: make mutex more efficient
- common: some async compression infrastructure (Haomai Wang)
- librados: add FULL\_TRY and FULL\_FORCE flags for dealing with full clusters or pools (Sage Weil)
- librados: fix notify completion race (#13114 Sage Weil)
- librados, libcephfs: randomize client nonces (Josh Durgin)

- librados: pybind: fix binary omap values (Robin H. Johnson)
- librbd: fix reads larger than the cache size (Lu Shi)
- librbd: metadata filter fixes (Haomai Wang)
- librbd: use write\_full when possible (Zhiqiang Wang)
- mds: avoid emitting cap warnigns before evicting session (John Spray)
- mds: fix expected holes in journal objects (#13167 Yan, Zheng)
- mds: fix SnapServer crash on deleted pool (John Spray)
- mds: many fixes (Yan, Zheng, John Spray, Greg Farnum)
- mon: add cache over MonitorDBStore (Kefu Chai)
- mon: ‘ceph osd metadata’ can dump all osds (Haomai Wang)
- mon: detect kv backend failures (Sage Weil)
- mon: fix CRUSH map test for new pools (Sage Weil)
- mon: fix min\_last\_epoch\_clean tracking (Kefu Chai)
- mon: misc scaling fixes (Sage Weil)
- mon: streamline session handling, fix memory leaks (Sage Weil)
- mon: upgrades must pass through hammer (Sage Weil)
- msg/async: many fixes (Haomai Wang)
- osd: cache proxy-write support (Zhiqiang Wang, Samuel Just)
- osd: configure promotion based on write recency (Zhiqiang Wang)
- osd: don’t send dup MMonGetOSDMap requests (Sage Weil, Kefu Chai)
- osd: erasure-code: fix SHEC floating point bug (#12936 Loic Dachary)
- osd: erasure-code: update to ISA-L 2.14 (Yuan Zhou)
- osd: fix hitset object naming to use GMT (Kefu Chai)
- osd: fix misc memory leaks (Sage Weil)
- osd: fix peek\_queue locking in FileStore (Xinze Chi)
- osd: fix promotion vs full cache tier (Samuel Just)
- osd: fix replay requeue when pg is still activating (#13116 Samuel Just)

- osd: fix scrub stat bugs (Sage Weil, Samuel Just)
- osd: force promotion for ops EC can't handle (Zhiqiang Wang)
- osd: improve behavior on machines with large memory pages (Steve Capper)
- osd: merge multiple setattr calls into a setattrs call (Xinxin Shu)
- osd: newstore prototype (Sage Weil)
- osd: ObjectStore internal API refactor (Sage Weil)
- osd: SHEC no longer experimental
- osd: throttle evict ops (Yunchuan Wen)
- osd: upgrades must pass through hammer (Sage Weil)
- osd: use SEEK\_HOLE / SEEK\_DATA for sparse copy (Xinxin Shu)
- rbd: rbd-replay-prep and rbd-replay improvements (Jason Dillaman)
- rgw: expose the number of unhealthy workers through admin socket (Guang Yang)
- rgw: fix casing of Content-Type header (Robin H. Johnson)
- rgw: fix decoding of X-Object-Manifest from GET on Swift DLO (Radoslaw Rzarzynski)
- rgw: fix sysvinit script
- rgw: fix sysvinit script w/ multiple instances (Sage Weil, Pavan Rallabhandi)
- rgw: improve handling of already removed buckets in expirer (Radoslaw Rzarzynski)
- rgw: log to /var/log/ceph instead of /var/log/radosgw
- rgw: rework X-Trans-Id header to be conform with Swift API (Radoslaw Rzarzynski)
- rgw: s3 encoding-type for get bucket (Jeff Weber)
- rgw: set max buckets per user in ceph.conf (Vikhyat Umrao)
- rgw: support for Swift expiration API (Radoslaw Rzarzynski, Yehuda Sadeh)
- rgw: user rm is idempotent (Orit Wasserman)
- selinux policy (Boris Ranto, Milan Broz)
- systemd: many fixes (Sage Weil, Owen Synge, Boris Ranto, Dan van der Ster)
- systemd: run daemons as user ceph

## v9.0.3

This is the second to last batch of development work for the Infernalis cycle. The most intrusive change is an internal (non user-visible) change to the OSD's ObjectStore interface. Many fixes and improvements elsewhere across RGW, RBD, and another big pile of CephFS scrub/repair improvements.

## Upgrading

---

- The return code for librbd's `rbd_aio_read` and `Image::aio_read` API methods no longer returns the number of bytes read upon success. Instead, it returns 0 upon success and a negative value upon failure.
- ‘ceph scrub’, ‘ceph compact’ and ‘ceph sync force’ are now deprecated. Users should instead use ‘ceph mon scrub’, ‘ceph mon compact’ and ‘ceph mon sync force’.
- ‘ceph mon\_metadata’ should now be used as ‘ceph mon metadata’.
- The `-dump-json` option of “`osdmaptool`” is replaced by `-dump json`.
- The commands of ‘`pg ls-by-{pool,primary,osd}`’ and ‘`pg ls`’ now take ‘recovering’ instead of ‘recovery’ to include the recovering pgs in the listed pgs.

## Notable Changes

---

- autotools: fix out of tree build (Krzysztof Kosinski)
- autotools: improve make check output (Loic Dachary)
- buffer: add invalidate\_crc() (Piotr Dalek)
- buffer: fix zero bug (#12252 Haomai Wang)
- build: fix junit detection on Fedora 22 (Ira Cooper)
- ceph-disk: install pip > 6.1 (#11952 Loic Dachary)
- cephfs-data-scan: many additions, improvements (John Spray)
- ceph: improve error output for ‘tell’ (#11101 Kefu Chai)
- ceph-objectstore-tool: misc improvements (David Zafman)
- ceph-objectstore-tool: refactoring and cleanup (John Spray)
- ceph\_test\_rados: test pipelined reads (Zhiqiang Wang)
- common: fix bit\_vector extent calc (#12611 Jason Dillaman)
- common: make work queue addition/removal thread safe (#12662 Jason Dillaman)
- common: optracker improvements (Zhiqiang Wang, Jianpeng Ma)
- crush: add `-check` to validate dangling names, max osd id (Kefu Chai)

- crush: cleanup, sync with kernel (Ilya Dryomov)
- crush: fix subtree base weight on adjust\_subtree\_weight (#11855 Sage Weil)
- crypto: fix NSS leak (Jason Dillaman)
- crypto: fix unbalanced init/shutdown (#12598 Zheng Yan)
- doc: misc updates (Kefu Chai, Owen Synge, Gael Fenet-Garde, Loic Dachary, Yannick Atchy-Dalama, Jiaying Ren, Kevin Caradant, Robert Maxime, Nicolas Yong, Germain Chipaux, Arthur Gorjux, Gabriel Sentucq, Clement Lebrun, Jean-Remi Deveaux, Clair Massot, Robin Tang, Thomas Laumondais, Jordan Dorne, Yuan Zhou, Valentin Thomas, Pierre Chaumont, Benjamin Troquereau, Benjamin Sesia, Vikhyat Umrao)
- erasure-code: cleanup (Kefu Chai)
- erasure-code: improve tests (Loic Dachary)
- erasure-code: shec: fix recovery bugs (Takanori Nakao, Shotaro Kawaguchi)
- libcephfs: add pread, pwrite (Jevon Qiao)
- libcephfs,ceph-fuse: cache cleanup (Zheng Yan)
- librados: add src\_fadvise\_flags for copy-from (Jianpeng Ma)
- librados: respect default\_crush\_ruleset on pool\_create (#11640 Yuan Zhou)
- librbd: fadvise for copy, export, import (Jianpeng Ma)
- librbd: handle NOCACHE fadvise flag (Jinapeng Ma)
- librbd: optionally disable allocation hint (Haomai Wang)
- librbd: prevent race between resize requests (#12664 Jason Dillaman)
- log: fix data corruption race resulting from log rotation (#12465 Samuel Just)
- mds: expose frags via asok (John Spray)
- mds: fix setting entire file layout in one setxattr (John Spray)
- mds: fix shutdown (John Spray)
- mds: handle misc corruption issues (John Spray)
- mds: misc fixes (Jianpeng Ma, Dan van der Ster, Zhang Zhi)
- mds: misc snap fixes (Zheng Yan)
- mds: store layout on header object (#4161 John Spray)
- misc performance and cleanup (Nathan Cutler, Xinxin Shu)
- mon: add NOFORWARD, OBSOLETE, DEPRECATE flags for mon commands (Joao Eduardo Luis)
- mon: add PG count to 'ceph osd df' output (Michal Jarzabek)
- mon: clean up, reorg some mon commands (Joao Eduardo Luis)
- mon: disallow >2 tiers (#11840 Kefu Chai)
- mon: fix log dump crash when debugging (Mykola Golub)
- mon: fix metadata update race (Mykola Golub)

- mon: fix refresh (#11470 Joao Eduardo Luis)
- mon: make blocked op messages more readable (Jianpeng Ma)
- mon: only send mon metadata to supporting peers (Sage Weil)
- mon: periodic background scrub (Joao Eduardo Luis)
- mon: prevent pgp\_num > pg\_num (#12025 Xinxin Shu)
- mon: reject large max\_mds values (#12222 John Spray)
- msgr: add ceph\_perf\_msgr tool (Hoamai Wang)
- msgr: async: fix seq handling (Hoamai Wang)
- msgr: xio: fastpath improvements (Raju Kurunkad)
- msgr: xio: sync with accellio v1.4 (Vu Pham)
- osd: clean up temp object if promotion fails (Jianpeng Ma)
- osd: constrain collections to meta and PGs (normal and temp) (Sage Weil)
- osd: filestore: clone using splice (Jianpeng Ma)
- osd: filestore: fix recursive lock (Xinxin Shu)
- osd: fix dup promotion lost op bug (Zhiqiang Wang)
- osd: fix temp-clearing (David Zafman)
- osd: include a temp namespace within each collection/pgid (Sage Weil)
- osd: low and high speed flush modes (Mingxin Liu)
- osd: peer\_features includes self (David Zafman)
- osd: recovery, peering fixes (#11687 Samuel Just)
- osd: require firefly features (David Zafman)
- osd: set initial crush weight with more precision (Sage Weil)
- osd: use a temp object for recovery (Sage Weil)
- osd: use blkid to collection partition information (Joseph Handzik)
- rados: add -striper option to use libradosstriper (#10759 Sebastien Ponce)
- radosgw-admin: fix subuser modify output (#12286 Guce)
- rados: handle -snapid arg properly (Abhishek Lekshmanan)
- rados: improve bench buffer handling, performance (Piotr Dalek)
- rados: new pool import implementation (John Spray)
- rbd: fix link issues (Jason Dillaman)
- rbd: improve CLI arg parsing, usage (Ilya Dryomov)
- rbd: recognize queue\_depth kernel option (Ilya Dryomov)

- rbd: support G and T units for CLI (Abhishek Lekshmanan)
- rbd: use image-spec and snap-spec in help (Vikhyat Umrao, Ilya Dryomov)
- rest-bench: misc fixes (Shawn Chen)
- rest-bench: support https (#3968 Yuan Zhou)
- rgw: add max multipart upload parts (#12146 Abhishek Dixit)
- rgw: add Transaction-Id to response (Abhishek Dixit)
- rgw: document layout of pools and objects (Pete Zaitcev)
- rgw: do not preserve ACLs when copying object (#12370 Yehuda Sadeh)
- rgw: fix Connection: header handling (#12298 Wido den Hollander)
- rgw: fix data corruptions race condition (#11749 Wuxingyi)
- rgw: fix JSON response when getting user quota (#12117 Wuxingyi)
- rgw: force content\_type for swift bucket stats requests (#12095 Orit Wasserman)
- rgw: improved support for swift account metadata (Radoslaw Zarzynski)
- rgw: make max put size configurable (#6999 Yuan Zhou)
- rgw: orphan detection tool (Yehuda Sadeh)
- rgw: swift: do not override sent content type (#12363 Orit Wasserman)
- rgw: swift: set Content-Length for account GET (#12158 Radoslav Zarzynski)
- rpm: always rebuild and install man pages for rpm (Owen Synge)
- rpm: misc fixes (Boris Ranto, Owen Synge, Ken Dreyer, Ira Cooper)
- systemd: logrotate fixes (Tim Seron, Lars Marowsky-Bree, Nathan Cutler)
- sysvinit compat: misc fixes (Owen Synge)
- test: misc fs test improvements (John Spray, Loic Dachary)
- test: python tests, linter cleanup (Alfredo Deza)

## v9.0.2

This development release features more of the OSD work queue unification, randomized osd scrub times, a huge pile of librbd fixes, more MDS repair and snapshot fixes, and a significant amount of work on the tests and build infrastructure.

## Notable Changes

- buffer: some cleanup (Michal Jarzabek)
- build: cmake: fix nss linking (Danny Al-Gaaf)

- build: cmake: misc fixes (Orit Wasserman, Casey Bodley)
- build: install-deps: misc fixes (Loic Dachary)
- build: make\_dist\_tarball.sh (Sage Weil)
- ceph-detect-init: added Linux Mint (Michal Jarzabek)
- ceph-detect-init: robust init system detection (Owen Synge, Loic Dachary)
- ceph-disk: ensure 'zap' only operates on a full disk (#11272 Loic Dachary)
- ceph-disk: misc fixes to respect init system (Loic Dachary, Owen Synge)
- ceph-disk: support NVMe device partitions (#11612 Ilja Slepnev)
- ceph: fix 'df' units (Zhe Zhang)
- ceph: fix parsing in interactive cli mode (#11279 Kefu Chai)
- ceph-objectstore-tool: many many changes (David Zafman)
- ceph-post-file: misc fixes (Joey McDonald, Sage Weil)
- client: avoid sending unnecessary FLUSHSNAP messages (Yan, Zheng)
- client: exclude setfilelock when calculating oldest tid (Yan, Zheng)
- client: fix error handling in check\_pool\_perm (John Spray)
- client: fsync waits only for inode's caps to flush (Yan, Zheng)
- client: invalidate kernel dcache when cache size exceeds limits (Yan, Zheng)
- client: make fsync wait for unsafe dir operations (Yan, Zheng)
- client: pin lookup dentry to avoid inode being freed (Yan, Zheng)
- common: detect overflow of int config values (#11484 Kefu Chai)
- common: fix json parsing of utf8 (#7387 Tim Serong)
- common: fix leak of pthread\_mutexattr (#11762 Ketor Meng)
- crush: respect default replicated ruleset config on map creation (Ilya Dryomov)
- deb, rpm: move ceph-objectstore-tool to ceph (Ken Dreyer)
- doc: man page updates (Kefu Chai)
- doc: misc updates (#11396 Nilamdyuti, Francois Lafont, Ken Dreyer, Kefu Chai)
- init-radosgw: merge with sysv version; fix enumeration (Sage Weil)

- librados: add config observer (Alistair Strachan)
- librbd: add const for single-client-only features (Josh Durgin)
- librbd: add deep-flatten operation (Jason Dillaman)
- librbd: avoid blocking aio API methods (#11056 Jason Dillaman)
- librbd: fix fast diff bugs (#11553 Jason Dillaman)
- librbd: fix image format detection (Zhiqiang Wang)
- librbd: fix lock ordering issue (#11577 Jason Dillaman)
- librbd: flatten/copyup fixes (Jason Dillaman)
- librbd: lockdep, helgrind validation (Jason Dillaman, Josh Durgin)
- librbd: only update image flags while hold exclusive lock (#11791 Jason Dillaman)
- librbd: return result code from close (#12069 Jason Dillaman)
- librbd: tolerate old osds when getting image metadata (#11549 Jason Dillaman)
- mds: do not add snapped items to bloom filter (Yan, Zheng)
- mds: fix handling for missing mydir dirfrag (#11641 John Spray)
- mds: fix rejoin (Yan, Zheng)
- mds: fix stra reintegration (Yan, Zheng)
- mds: fix suicide beason (John Spray)
- mds: misc repair improvements (John Spray)
- mds: misc snapshot fixes (Yan, Zheng)
- mds: respawn instead of suicide on blacklist (John Spray)
- misc coverity fixes (Danny Al-Gaaf)
- mon: add ‘mon\_metadata <id>’ command (Kefu Chai)
- mon: add ‘node ls ...’ command (Kefu Chai)
- mon: disallow ec pools as tiers (#11650 Samuel Just)
- mon: fix mds beacon replies (#11590 Kefu Chai)
- mon: fix ‘pg ls’ sort order, state names (#11569 Kefu Chai)
- mon: normalize erasure-code profile for storage and comparison (Loic Dachary)

- mon: optionally specify osd id on 'osd create' (Mykola Golub)
- mon: 'osd tree' fixes (Kefu Chai)
- mon: prevent pool with snapshot state from being used as a tier (#11493 Sage Weil)
- mon: refine check\_remove\_tier checks (#11504 John Spray)
- mon: remove spurious who arg from 'mds rm ...' (John Spray)
- msgr: async: misc fixes (Haomai Wang)
- msgr: xio: fix ip and nonce (Raju Kurunkad)
- msgr: xio: improve lane assignment (Vu Pham)
- msgr: xio: misc fixes (Vu Pham, Cosey Bodley)
- osd: avoid transaction append in some cases (Sage Weil)
- osdc/Objecter: allow per-pool calls to op\_cancel\_writes (John Spray)
- osd: elminiate txn apend, ECSubWrite copy (Samuel Just)
- osd: filejournal: cleanup (David Zafman)
- osd: fix check\_for\_full (Henry Chang)
- osd: fix dirty accounting in make\_writeable (Zhiqiang Wang)
- osd: fix osdmap dump of blacklist items (John Spray)
- osd: fix snap flushing from cache tier (again) (#11787 Samuel Just)
- osd: fix snap handling on promotion (#11296 Sam Just)
- osd: handle log split with overlapping entries (#11358 Samuel Just)
- osd: keyvaluestore: misc fixes (Varada Kari)
- osd: make suicide timeouts individually configurable (Samuel Just)
- osd: move scrub in OpWQ (Samuel Just)
- osd: pool size change triggers new interval (#11771 Samuel Just)
- osd: randomize scrub times (#10973 Kefu Chai)
- osd: refactor scrub and digest recording (Sage Weil)
- osd: refuse first write to EC object at non-zero offset (Jianpeng Ma)
- osd: stripe over small xattrs to fit in XFS's 255 byte inline limit (Sage Weil,

Ning Yao)

- osd: sync object\_map on syncfs (Samuel Just)
- osd: take excl lock of op is rw (Samuel Just)
- osd: WBThrottle cleanups (Jianpeng Ma)
- pycephfs: many fixes for bindings (Haomai Wang)
- rados: bench: add -no-verify option to improve performance (Piotr Dalek)
- rados: misc bench fixes (Dmitry Yatsushkevich)
- rbd: add disk usage tool (#7746 Jason Dillaman)
- rgw: always check if token is expired (#11367 Anton Aksola, Riku Lehto)
- rgw: conversion tool to repair broken multipart objects (#12079 Yehuda Sadeh)
- rgw: do not enclose bucket header in quotes (#11860 Wido den Hollander)
- rgw: error out if frontend did not send all data (#11851 Yehuda Sadeh)
- rgw: fix assignment of copy obj attributes (#11563 Yehuda Sadeh)
- rgw: fix reset\_loc (#11974 Yehuda Sadeh)
- rgw: improve content-length env var handling (#11419 Robin H. Johnson)
- rgw: only scan for objects not in a namespace (#11984 Yehuda Sadeh)
- rgw: remove trailing :port from HTTP\_HOST header (Sage Weil)
- rgw: shard work over multiple librados instances (Pavan Rallabhandi)
- rgw: swift: enforce Content-Type in response (#12157 Radoslaw Zarzynski)
- rgw: use attrs from source bucket on copy (#11639 Javier M. Mellid)
- rocksdb: pass options as single string (Xiaoxi Chen)
- rpm: many spec file fixes (Owen Synge, Ken Dreyer)
- tests: fixes for rbd xstests (Douglas Fuller)
- tests: fix tiering health checks (Loic Dachary)
- tests for low-level performance (Haomai Wang)
- tests: many ec non-regression improvements (Loic Dachary)
- tests: many many ec test improvements (Loic Dachary)

- upstart: throttle restarts (#11798 Sage Weil, Greg Farnum)

## v9.0.1

---

This development release is delayed a bit due to tooling changes in the build environment. As a result the next one (v9.0.2) will have a bit more work than is usual.

Highlights here include lots of RGW Swift fixes, RBD feature work surrounding the new object map feature, more CephFS snapshot fixes, and a few important CRUSH fixes.

## Notable Changes

---

- auth: cache/reuse crypto lib key objects, optimize msg signature check (Sage Weil)
- build: allow tcmalloc-minimal (Thorsten Behrens)
- build: do not build ceph-dencoder with tcmalloc (#10691 Boris Ranto)
- build: fix pg ref disabling (William A. Kennington III)
- build: install-deps.sh improvements (Loic Dachary)
- build: misc fixes (Boris Ranto, Ken Dreyer, Owen Synge)
- ceph-authtool: fix return code on error (Gerhard Muntingh)
- ceph-disk: fix zap sgdisk invocation (Owen Synge, Thorsten Behrens)
- ceph-disk: pass -cluster arg on prepare subcommand (Kefu Chai)
- ceph-fuse, libcephfs: drop inode when rmdir finishes (#11339 Yan, Zheng)
- ceph-fuse, libcephfs: fix uninlne (#11356 Yan, Zheng)
- ceph-monstore-tool: fix store-copy (Huangjun)
- common: add perf counter descriptions (Alyona Kiseleva)
- common: fix throttle max change (Henry Chang)
- crush: fix crash from invalid 'take' argument (#11602 Shiva Rkreddy, Sage Weil)
- crush: fix divide-by-2 in straw2 (#11357 Yann Dupont, Sage Weil)
- deb: fix rest-bench-dbg and ceph-test-dbg dependendies (Ken Dreyer)
- doc: document region hostnames (Robin H. Johnson)
- doc: update release schedule docs (Loic Dachary)

- init-radosgw: run radosgw as root (#11453 Ken Dreyer)
- librados: fadvise flags per op (Jianpeng Ma)
- librbd: allow additional metadata to be stored with the image (Haomai Wang)
- librbd: better handling for dup flatten requests (#11370 Jason Dillaman)
- librbd: cancel in-flight ops on watch error (#11363 Jason Dillaman)
- librbd: default new images to format 2 (#11348 Jason Dillaman)
- librbd: fast diff implementation that leverages object map (Jason Dillaman)
- librbd: fix snapshot creation when other snap is active (#11475 Jason Dillaman)
- librbd: new diff\_iterate2 API (Jason Dillaman)
- librbd: object map rebuild support (Jason Dillaman)
- logrotate.d: prefer service over invoke-rc.d (#11330 Win Hierman, Sage Weil)
- mds: avoid getting stuck in XLOCKDONE (#11254 Yan, Zheng)
- mds: fix integer truncation on large client ids (Henry Chang)
- mds: many snapshot and stray fixes (Yan, Zheng)
- mds: persist completed\_requests reliably (#11048 John Spray)
- mds: separate safe\_pos in Journaler (#10368 John Spray)
- mds: snapshot rename support (#3645 Yan, Zheng)
- mds: warn when clients fail to advance oldest\_client\_tid (#10657 Yan, Zheng)
- misc cleanups and fixes (Danny Al-Gaaf)
- mon: fix average utilization calc for 'osd df' (Mykola Golub)
- mon: fix variance calc in 'osd df' (Sage Weil)
- mon: improve callout to crushtool (Mykola Golub)
- mon: prevent bucket deletion when referenced by a crush rule (#11602 Sage Weil)
- mon: prime pg\_temp when CRUSH map changes (Sage Weil)
- monclient: flush\_log (John Spray)
- msgr: async: many many fixes (Haomai Wang)
- msgr: simple: fix clear\_pipe (#11381 Haomai Wang)

- osd: add latency perf counters for tier operations (Xinze Chi)
- osd: avoid multiple hit set insertions (Zhiqiang Wang)
- osd: break PG removal into multiple iterations (#10198 Guang Yang)
- osd: check scrub state when handling map (Jianpeng Ma)
- osd: fix endless repair when object is unrecoverable (Jianpeng Ma, Kefu Chai)
- osd: fix pg resurrection (#11429 Samuel Just)
- osd: ignore non-existent osds in unfound calc (#10976 Mykola Golub)
- osd: increase default max open files (Owen Synge)
- osd: prepopulate needs\_recovery\_map when only one peer has missing (#9558 Guang Yang)
- osd: relax reply order on proxy read (#11211 Zhiqiang Wang)
- osd: skip promotion for flush/evict op (Zhiqiang Wang)
- osd: write journal header on clean shutdown (Xinze Chi)
- qa: run-make-check.sh script (Loic Dachary)
- rados bench: misc fixes (Dmitry Yatsushkevich)
- rados: fix error message on failed pool removal (Wido den Hollander)
- radosgw-admin: add ‘bucket check’ function to repair bucket index (Yehuda Sadeh)
- rbd: allow unmapping by spec (Ilya Dryomov)
- rbd: deprecate -new-format option (Jason Dillman)
- rgw: do not set content-type if length is 0 (#11091 Orit Wasserman)
- rgw: don’t use end\_marker for namespaced object listing (#11437 Yehuda Sadeh)
- rgw: fail if parts not specified on multipart upload (#11435 Yehuda Sadeh)
- rgw: fix GET on swift account when limit == 0 (#10683 Radoslaw Zarzynski)
- rgw: fix broken stats in container listing (#11285 Radoslaw Zarzynski)
- rgw: fix bug in domain/subdomain splitting (Robin H. Johnson)
- rgw: fix civetweb max threads (#10243 Yehuda Sadeh)
- rgw: fix copy metadata, support X-Copied-From for swift (#10663 Radoslaw Zarzynski)

- rgw: fix locator for objects starting with \_ (#11442 Yehuda Sadeh)
- rgw: fix multipart upload in retry path (#11604 Yehuda Sadeh)
- rgw: fix quota enforcement on POST (#11323 Sergey Arkhipov)
- rgw: fix return code on missing upload (#11436 Yehuda Sadeh)
- rgw: force content type header on responses with no body (#11438 Orit Wasserman)
- rgw: generate new object tag when setting attrs (#11256 Yehuda Sadeh)
- rgw: issue aio for first chunk before flush cached data (#11322 Guang Yang)
- rgw: make read user buckets backward compat (#10683 Radoslaw Zarzynski)
- rgw: merge manifests properly with prefix override (#11622 Yehuda Sadeh)
- rgw: return 412 on bad limit when listing buckets (#11613 Yehuda Sadeh)
- rgw: send ETag, Last-Modified for swift (#11087 Radoslaw Zarzynski)
- rgw: set content length on container GET, PUT, DELETE, HEAD (#10971, #11036 Radoslaw Zarzynski)
- rgw: support end marker on swift container GET (#10682 Radoslaw Zarzynski)
- rgw: swift: fix account listing (#11501 Radoslaw Zarzynski)
- rgw: swift: set content-length on keystone tokens (#11473 Herv Rousseau)
- rgw: use correct oid for gc chains (#11447 Yehuda Sadeh)
- rgw: use unique request id for civetweb (#10295 Orit Wasserman)
- rocksdb, leveldb: fix compact\_on\_mount (Xiaoxi Chen)
- rocksdb: add perf counters for get/put latency (Xinxin Shu)
- rpm: add suse firewall files (Tim Serong)
- rpm: misc systemd and suse fixes (Owen Synge, Nathan Cutler)

## v9.0.0

This is the first development release for the Infernalis cycle, and the first Ceph release to sport a version number from the new numbering scheme. The “9” indicates this is the 9th release cycle-I (for Infernalis) is the 9th letter. The first “0” indicates this is a development release (“1” will mean release candidate and “2” will mean stable release), and the final “0” indicates this is the first such development release.

A few highlights include:

- a new ‘ceph daemonperf’ command to watch perfcounter stats in realtime
- reduced MDS memory usage
- many MDS snapshot fixes
- librbd can now store options in the image itself
- many fixes for RGW Swift API support
- OSD performance improvements
- many doc updates and misc bug fixes

## Notable Changes

---

- aarch64: add optimized version of crc32c (Yazen Ghannam, Steve Capper)
- auth: reinit NSS after fork() (#11128 Yan, Zheng)
- build: disable LTTNG by default (#11333 Josh Durgin)
- build: fix ppc build (James Page)
- build: install-deps: support OpenSUSE (Loic Dachary)
- build: misc cmake fixes (Matt Benjamin)
- ceph-disk: follow ceph-osd hints when creating journal (#9580 Sage Weil)
- ceph-disk: handle re-using existing partition (#10987 Loic Dachary)
- ceph-disk: improve parted output parsing (#10983 Loic Dachary)
- ceph-disk: make suppression work for activate-all and activate-journal (Dan van der Ster)
- ceph-disk: misc fixes (Alfredo Deza)
- ceph-fuse, libcephfs: don’t clear COMPLETE when trimming null (Yan, Zheng)
- ceph-fuse, libcephfs: hold exclusive caps on dirs we “own” (#11226 Greg Farnum)
- ceph-fuse: do not require successful remount when unmounting (#10982 Greg Farnum)
- ceph: new ‘ceph daemonperf’ command (John Spray, Mykola Golub)
- common: PriorityQueue tests (Kefu Chai)
- common: add descriptions to perfcounters (Kiseleva Alyona)

- common: fix LTTNG vs fork issue (Josh Durgin)
- crush: fix has\_v4\_buckets (#11364 Sage Weil)
- crushtool: fix order of operations, usage (Sage Weil)
- debian: minor package reorg (Ken Dreyer)
- doc: docuemnt object corpus generation (#11099 Alexis Normand)
- doc: fix gender neutrality (Alexandre Maragone)
- doc: fix install doc (#10957 Kefu Chai)
- doc: fix sphinx issues (Kefu Chai)
- doc: mds data structure docs (Yan, Zheng)
- doc: misc updates (Nilamdyuti Goswami, Vartika Rai, Florian Haas, Loic Dachary, Simon Guinot, Andy Allan, Alistair Israel, Ken Dreyer, Robin Rehu, Lee Revell, Florian Marsylle, Thomas Johnson, Bosse Klykken, Travis Rhoden, Ian Kelling)
- doc: swift tempurls (#10184 Abhishek Lekshmanan)
- doc: switch doxygen integration back to breathe (#6115 Kefu Chai)
- erasure-code: update ISA-L to 2.13 (Yuan Zhou)
- gmock: switch to submodule (Danny Al-Gaaf, Loic Dachary)
- hadoop: add terasort test (Noah Watkins)
- java: fix libcephfs bindings (Noah Watkins)
- libcephfs,ceph-fuse: fix request resend on cap reconnect (#10912 Yan, Zheng)
- librados: define C++ flags from C constants (Josh Durgin)
- librados: fix last\_force\_resent handling (#11026 Jianpeng Ma)
- librados: fix memory leak from C\_TwoContexts (Xiong Yiliang)
- librados: fix striper when stripe\_count = 1 and stripe\_unit != object\_size (#11120 Yan, Zheng)
- librados: op perf counters (John Spray)
- librados: pybind: fix write() method return code (Javier Guerra)
- libradosstriper: fix leak (Danny Al-Gaaf)
- librbd: add purge\_on\_error cache behavior (Jianpeng Ma)
- librbd: misc aio fixes (#5488 Jason Dillaman)

- librbd: misc rbd fixes (#11478 #11113 #11342 #11380 Jason Dillaman, Zhiqiang Wang)
- librbd: readahead fixes (Zhiqiang Wang)
- librbd: store metadata, including config options, in image (Haomai Wang)
- mds: add ‘damaged’ state to MDSMap (John Spray)
- mds: add nicknames for perfcounters (John Spray)
- mds: disable problematic rstat propagation into snap parents (Yan, Zheng)
- mds: fix mydir replica issue with shutdown (#10743 John Spray)
- mds: fix out-of-order messages (#11258 Yan, Zheng)
- mds: fix shutdown with strays (#10744 John Spray)
- mds: fix snapshot fixes (Yan, Zheng)
- mds: fix stray handling (John Spray)
- mds: flush immediately in do\_open\_truncate (#11011 John Spray)
- mds: improve dump methods (John Spray)
- mds: misc journal cleanups and fixes (#10368 John Spray)
- mds: new SessionMap storage using omap (#10649 John Spray)
- mds: reduce memory consumption (Yan, Zheng)
- mds: throttle purge stray operations (#10390 John Spray)
- mds: tolerate clock jumping backwards (#11053 Yan, Zheng)
- misc coverity fixes (Danny Al-Gaaf)
- mon: do not deactivate last mds (#10862 John Spray)
- mon: make osd get pool ‘all’ only return applicable fields (#10891 Michal Jarzabek)
- mon: warn on bogus cache tier config (Jianpeng Ma)
- msg/async: misc bug fixes and updates (Haomai Wang)
- msg/simple: fix connect\_seq assert (Haomai Wang)
- msg/xio: misc fixes (#10735 Matt Benjamin, Kefu Chai, Danny Al-Gaaf, Raju Kurunkad, Vu Pham)
- msg: unit tests (Haomai Wang)

- objectcacher: misc bug fixes (Jianpeng Ma)
- os/filestore: enlarge getxattr buffer size (Jianpeng Ma)
- osd: EIO injection (David Zhang)
- osd: add misc perfcounters (Xinze Chi)
- osd: add simple sleep injection in recovery (Sage Weil)
- osd: allow SEEK\_HOLE/SEEK\_DATA for sparse read (Zhiqiang Wang)
- osd: avoid dup omap sets for in pg metadata (Sage Weil)
- osd: clean up some constness, privateness (Kefu Chai)
- osd: erasure-code: drop entries according to LRU (Andreas-Joachim Peters)
- osd: fix negative degraded stats during backfill (Guang Yang)
- osd: misc fixes (Ning Yao, Kefu Chai, Xinze Chi, Zhiqiang Wang, Jianpeng Ma)
- pybind: pep8 cleanups (Danny Al-Gaaf)
- qa: fix filelock\_interrupt.py test (Yan, Zheng)
- qa: improve ceph-disk tests (Loic Dachary)
- qa: improve docker build layers (Loic Dachary)
- rados: translate erno to string in CLI (#10877 Kefu Chai)
- rbd: accept map options config option (Ilya Dryomov)
- rbd: cli: fix arg parsing with -io-pattern (Dmitry Yatsushkevich)
- rbd: fix error messages (#2862 Rajesh Nambiar)
- rbd: update rbd man page (Ilya Dryomov)
- rbd: update xfstests tests (Douglas Fuller)
- rgw: add X-Timestamp for Swift containers (#10938 Radoslaw Zarzynski)
- rgw: add missing headers to Swift container details (#10666 Ahmad Faheem, Dmytro Iurchenko)
- rgw: add stats to headers for account GET (#10684 Yuan Zhou)
- rgw: do not prefetch data for HEAD requests (Guang Yang)
- rgw: don't clobber bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: don't use rgw\_socket\_path if frontend is configured (#11160 Yehuda Sadeh)

- rgw: enforce Content-Lenth for POST on Swift cont/obj (#10661 Radoslaw Zarzynski)
- rgw: fix handling empty metadata items on Swift container (#11088 Radoslaw Zarzynski)
- rgw: fix log rotation (Wuxingyi)
- rgw: generate Date header for civetweb (#10873 Radoslaw Zarzynski)
- rgw: make init script wait for radosgw to stop (#11140 Dmitry Yatsushkevich)
- rgw: make quota/gc threads configurable (#11047 Guang Yang)
- rgw: pass in civetweb configurables (#10907 Yehuda Sadeh)
- rgw: rectify 202 Accepted in PUT response (#11148 Radoslaw Zarzynski)
- rgw: remove meta file after deleting bucket (#11149 Orit Wasserman)
- rgw: swift: allow setting attributes with COPY (#10662 Ahmad Faheem, Dmytro Iurchenko)
- rgw: swift: fix metadata handling on copy (#10645 Radoslaw Zarzynski)
- rgw: swift: send Last-Modified header (#10650 Radoslaw Zarzynski)
- rgw: update keystone cache with token info (#11125 Yehuda Sadeh)
- rgw: update to latest civetweb, enable config for IPv6 (#10965 Yehuda Sadeh)
- rocksdb: update to latest (Xiaoxi Chen)
- rpm: loosen ceph-test dependencies (Ken Dreyer)

# v0.94.10 Hammer

---

This Hammer point release fixes several bugs and adds two new features.

We recommend that all hammer v0.94.x users upgrade.

For more detailed information, see [the complete changelog](#).

## New Features

---

ceph-objectstore-tool and ceph-monstore-tool now enable user to rebuild the monitor database from OSDs. (This feature is especially useful when all monitors fail to boot due to leveldb corruption.)

In RADOS Gateway, it is now possible to reshards an existing bucket's index using an off-line tool.

Usage:

```
$ radosgw-admin bucket reshards -bucket=<bucket_name> -num_shards=<num_shards>
```

This will create a new linked bucket instance that points to the newly created index objects. The old bucket instance still exists and currently it's up to the user to manually remove the old bucket index objects. (Note that bucket resharding currently requires that all IO (especially writes) to the specific bucket is quiesced.)

## Other Notable Changes

---

- build/ops: ceph-create-keys loops forever ([issue#17753](#), [pr#12805](#), Alfredo Deza)
- build/ops: improve ceph.in error message ([issue#11101](#), [pr#10905](#), Kefu Chai)
- build/ops: make stop.sh more portable ([issue#16918](#), [pr#10569](#), Mykola Golub)
- build/ops: remove SYSTEMD\_RUN from initscript ([issue#16440](#), [issue#7627](#), [pr#9873](#), Vladislav Odintsov)
- cephx: Fix multiple segfaults due to attempts to encrypt or decrypt ([issue#16266](#), [pr#11930](#), Brad Hubbard)
- common: SIGABRT in TrackedOp::dump() via dump\_ops\_in\_flight() ([issue#8885](#), [pr#12121](#), Jianpeng Ma, Zhiqiang Wang, David Zafman)
- common: os/ObjectStore: fix \_update\_op for split dest\_cid ([issue#15345](#), [pr#12071](#), Sage Weil)
- crush: reset bucket->h.items[i] when removing tree item ([issue#16525](#), [pr#10724](#),

Kefu Chai)

- doc: add “Upgrading to Hammer” section ([issue#17386](#), [pr#11372](#), Kefu Chai)
- doc: add orphan options to radosgw-admin -help and man page ([issue#17281](#), [issue#17280](#), [pr#11140](#), Abhishek Lekshmanan, Casey Bodley, Ken Dreyer, Thomas Serlin)
- doc: clarify that RGW bucket object versioning is supported ([issue#16574](#), [pr#10437](#), Yuan Zhou, shawn chen)
- librados: bad flags can crash the osd ([issue#16012](#), [pr#11936](#), Jianpeng Ma, Sage Weil)
- librbd: ceph 10.2.2 rbd status on image format 2 returns “(2) No such file or directory” ([issue#16887](#), [pr#10987](#), Jason Dillaman)
- librbd: diffs to clone’s first snapshot should include parent diffs ([issue#18068](#), [pr#12446](#), Jason Dillaman)
- librbd: image.stat() call in librbdpy fails sometimes ([issue#17310](#), [pr#11949](#), Jason Dillaman)
- librbd: request exclusive lock if current owner cannot execute op ([issue#16171](#), [pr#12018](#), Mykola Golub)
- mds: fix cephfs-java ftruncate unit test failure ([issue#11258](#), [pr#11939](#), Yan, Zheng)
- mon: %USED of ceph df is wrong ([issue#16933](#), [pr#11934](#), Kefu Chai)
- mon: MonmapMonitor should return success when MON will be removed ([issue#17725](#), [pr#12006](#), Joao Eduardo Luis)
- mon: OSDMonitor: Missing nearfull flag set ([issue#17390](#), [pr#11273](#), Igor Podoski)
- mon: OSDs marked OUT wrongly after monitor failover ([issue#17719](#), [pr#11946](#), Dong Wu)
- mon: fix memory leak in prepare\_beacon ([issue#17285](#), [pr#10238](#), Igor Podoski)
- mon: osd flag health message is misleading ([issue#18175](#), [pr#12687](#), Sage Weil)
- mon: prepare\_pgtemp needs to only update up\_thru if newer than the existing one ([issue#16185](#), [pr#11937](#), Samuel Just)
- mon: return size\_t from MonitorDBStore::Transaction::size() ([issue#14217](#), [pr#10904](#), Kefu Chai)
- mon: send updated monmap to its subscribers ([issue#17558](#), [pr#11457](#), Kefu Chai)
- msgr: OpTracker needs to release the message throttle in \_unregistered

- ([issue#14248](#), [pr#11938](#), Samuel Just)
- msgr: simple/Pipe: error decoding addr ([issue#18072](#), [pr#12266](#), Sage Weil)
  - osd: PG::\_update\_calc\_stats wrong for CRUSH\_ITEM\_NONE up set items ([issue#16998](#), [pr#11933](#), Samuel Just)
  - osd: PG::choose\_acting valgrind error or ./common/hobject.h: 182: FAILED assert(!max || (\*this == hobject\_t(hobject\_t::get\_max()))) ([issue#13967](#), [pr#11932](#), Tao Chang)
  - osd: ReplicatedBackend::build\_push\_op: add a second config to limit omap entries/chunk independently of object data ([issue#16128](#), [pr#12417](#), Wanlong Gao)
  - osd: crash on EIO during deep-scrubbing ([issue#16034](#), [pr#11935](#), Nathan Cutler)
  - osd: filestore: FALLOC\_FL\_PUNCH\_HOLE must be used with FALLOC\_FL\_KEEP\_SIZE ([issue#18446](#), [pr#13041](#), xinxin shu)
  - osd: fix cached\_removed\_snaps bug in PGPool::update after map gap ([issue#18628](#), [issue#15943](#), [pr#12906](#), Samuel Just)
  - osd: fix collection\_list shadow return value ([issue#17713](#), [pr#11927](#), Haomai Wang)
  - osd: fix fiemap issue in xfs when #extents > 1364 ([issue#17610](#), [pr#11615](#), Kefu Chai, Ning Yao)
  - osd: update PGPool to detect map gaps and reset cached\_removed\_snaps ([issue#15943](#), [pr#11676](#), Samuel Just)
  - rbd: export diff should open image as read-only ([issue#17671](#), [pr#11948](#), liyankun)
  - rbd: fix parameter check ([issue#18237](#), [pr#12312](#), Yankun Li)
  - rbd: fix possible rbd data corruption ([issue#16002](#), [pr#11618](#), Yan, Zheng, Greg Farnum)
  - rgw: Anonymous user is able to read bucket with authenticated read ACL ([issue#13207](#), [pr#11045](#), rahul.1aggarwal@gmail.com)
  - rgw: COPY broke multipart files uploaded under dumpling ([issue#16435](#), [pr#11950](#), Yehuda Sadeh)
  - rgw: TempURL in radosgw behaves now like its Swift's counterpart. ([issue#18316](#), [pr#12619](#), Radoslaw Zarzynski)
  - rgw: default quota fixes ([issue#16410](#), [pr#10839](#), Pavan Rallabhandi, Daniel Grynewicz)
  - rgw: do not abort when accept a CORS request with short origin ([issue#18187](#), [pr#12398](#), LiuYang)

- rgw: do not omap\_getvals with (u64)-1 max ([issue#17985](#), [pr#12418](#), Yehuda Sadeh, Sage Weil)
- rgw: fix crash when client posts object with null condition ([issue#17635](#), [pr#11809](#), Yehuda Sadeh)
- rgw: fix inconsistent uid/email handling in radosgw-admin ([issue#13598](#), [pr#11952](#), Matt Benjamin)
- rgw: implement offline resharding command ([issue#17745](#), [pr#12227](#), Yehuda Sadeh, Orit Wasserman, weiqiaomiao)
- rgw: swift: ranged request on a DLO provides wrong values in Content-Range HTTP header ([issue#13452](#), [pr#11951](#), Radoslaw Zarzynski)
- rgw: the value of total\_time is wrong in the result of 'radosgw-admin log show' opt ([issue#17598](#), [pr#11899](#), weiqiaomiao)
- tests: Cannot clone ceph/s3-tests.git (missing branch) ([issue#18384](#), [pr#12744](#), Orit Wasserman)
- tests: Cannot reserve CentOS 7.2 smithi machines ([issue#18401](#), [pr#12762](#), Nathan Cutler)
- tests: OSDs commit suicide in rbd suite when testing on btrfs ([issue#18397](#), [pr#12758](#), Nathan Cutler)
- tests: Workunits needlessly wget from git.ceph.com ([issue#18336](#), [issue#18271](#), [issue#18388](#), [pr#12685](#), Sage Weil, Nathan Cutler)
- tests: cephfs test failures (ceph.com/qa is broken, should be download.ceph.com/qa) ([issue#18574](#), [pr#13022](#), John Spray)
- tests: merge ceph-qa-suite ([pr#12455](#), Sage Weil)
- tests: objecter\_requests workunit fails on wip branches ([issue#18393](#), [pr#12759](#), Sage Weil)
- tests: populate mnt\_point in qa/tasks/ceph.py ([issue#18383](#), [pr#12743](#), Nathan Cutler)
- tests: qemu/tests/qemu-iotests/077 fails in dumpling, hammer, and jewel ([issue#10773](#), [pr#12423](#), Jason Dillaman)
- tests: run fs/thrash on xfs instead of btrfs ([issue#17151](#), [pr#13039](#), Nathan Cutler)
- tests: update Ubuntu image url after ceph.com refactor ([issue#18542](#), [pr#12957](#), Jason Dillaman)
- tests: update rbd/singleton/all/formatted-output.yaml to support ceph-ci \*

- ([issue#18440](#), [pr#12824](#) \*, Venky Shankar, Nathan Cutler)
- tools: add a tool to rebuild mon store from OSD ([issue#17179](#), [issue#17400](#), [pr#11125](#), Kefu Chai, xie xingguo)
  - tools: ceph-objectstore-tool crashes if -journal-path <a-directory> ([issue#17307](#), [pr#11929](#), Kefu Chai)
  - tools: ceph-objectstore-tool: add a way to split filestore directories offline ([issue#17220](#), [pr#11253](#), Josh Durgin)
  - tools: crushtool -compile generates output despite missing item ([issue#17306](#), [pr#11931](#), Kefu Chai)

## v0.94.9 Hammer

---

This Hammer point release fixes a build issue present in 0.94.8 that prevented us from generating packages for Ubuntu Precise and CentOS 6.x.

We recommend all users of v0.94.7 or older upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build/ops: revert: boost uuid makes valgrind complain ([pr#10913](#), Sage Weil)

## v0.94.8 Hammer

---

This Hammer point release fixes several bugs.

We recommend that all hammer v0.94.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build/ops: rocksdb do not link against tcmalloc if it's disabled ([issue#14799](#), [pr#10750](#), Sage Weil, Kefu Chai)
- build/ops: Add -D\_LARGEFILE64\_SOURCE to Linux build. ([issue#16611](#), [pr#10182](#), Ira Cooper)
- build/ops: boost uuid makes valgrind complain ([issue#12736](#), [pr#9741](#), Sage Weil, Rohan Mars)
- build/ops: ceph-disk s/by-parttype-uuid/by-parttypeuuid/ ([issue#15867](#), [pr#9107](#),

Nathan Cutler)

- common: add units to rados bench output and clean up formatting ([issue#12248](#), [pr#8960](#), Dmitry Yatsushkevich, Brad Hubbard, Gu Zhongyan)
- common: config set with negative value results in “error setting ‘filestore\_merge\_threshold’ to ‘-40’: (22) Invalid argument” ([issue#13829](#), [pr#10291](#), Brad Hubbard, Kefu Chai)
- common: linking to -lrbd causes process startup times to balloon ([issue#15225](#), [pr#8538](#), Richard W.M. Jones)
- doc: fix by-parttypeuuid in ceph-disk(8) nroff ([issue#15867](#), [pr#10699](#), Ken Dreyer)
- fs: double decreased the count to trim caps which will cause failing to respond to cache pressure ([issue#14319](#), [pr#8804](#), Zhi Zhang)
- log: do not repeat errors to stderr ([issue#14616](#), [pr#10227](#), Sage Weil)
- mds: failing file operations on kernel based cephfs mount point leaves unaccessible file behind on hammer 0.94.7 ([issue#16013](#), [pr#10198](#), Yan, Zheng)
- mds: fix stray purging in ‘stripe\_count > 1’ case ([issue#15050](#), [pr#8042](#), Yan, Zheng)
- mds: wrongly treat symlink inode as normal file/dir when symlink inode is stale on kcephfs ([issue#15702](#), [pr#9404](#), Zhi Zhang)
- mon: LibRadosMiscConnectFailure.ConnectFailure (not so intermittent) failure in upgrade/hammer-x ([issue#13992](#), [pr#8806](#), Sage Weil)
- mon: Monitor: validate prefix on handle\_command() ([issue#16297](#), [pr#10038](#), You Ji)
- mon: drop pg temps from not the current primary in OSDMonitor ([issue#16127](#), [pr#9893](#), Samuel Just)
- mon: fix calculation of %USED ([issue#15641](#), [pr#9125](#), Ruifeng Yang, David Zafman)
- mon: improve reweight\_by\_utilization() logic ([issue#15686](#), [pr#9416](#), xie xingguo)
- mon: pool quota alarm is not in effect ([issue#15478](#), [pr#8593](#), Danny Al-Gaaf)
- mon: wrong ceph get mdsmap assertion ([issue#14681](#), [pr#7542](#), Vicente Cheng)
- msgr: ceph-osd valgrind invalid reads/writes ([issue#15870](#), [pr#9238](#), Samuel Just)
- objecter: LibRadosWatchNotifyPPTests/LibRadosWatchNotifyPP.WatchNotify2Timeout/1 segv ([issue#15760](#), [pr#9400](#), Sage Weil)
- osd: OSD reporting ENOTEMPTY and crashing ([issue#14766](#), [pr#9277](#), Samuel Just)

- osd: When generating past intervals due to an import end at pg epoch and fix build\_past\_intervals\_parallel ([issue#12387](#), [issue#14438](#), [pr#8464](#), David Zafman)
- osd: acting\_primary not updated on split ([issue#15523](#), [pr#9001](#), Sage Weil)
- osd: assert(!actingbackfill.empty()): old watch timeout tries to queue repop on replica ([issue#15391](#), [pr#8665](#), Sage Weil)
- osd: assert(rollback\_info\_trimmed\_to == head) in PGLog ([issue#13965](#), [pr#8849](#), Samuel Just)
- osd: delete one of the repeated op->mark\_started in ReplicatedBackend::sub\_op\_modify\_impl ([issue#16572](#), [pr#9977](#), shun-s)
- osd: fix omap digest compare when scrub ([issue#16000](#), [pr#9271](#), Xinze Chi)
- osd: is\_split crash in handle\_pg\_create ([issue#15426](#), [pr#8805](#), Kefu Chai)
- osd: objects unfound after repair (fixed by repeering the pg) ([issue#15006](#), [pr#7961](#), Jianpeng Ma, Loic Dachary, Kefu Chai)
- osd: rados cppool omap to ec pool crashes osd ([issue#14695](#), [pr#8845](#), Jianpeng Ma)
- osd: remove all stale osdmmaps in handle\_osd\_map() ([issue#13990](#), [pr#9090](#), Kefu Chai)
- osd: send write and read sub ops on behalf of client ops at normal priority in ECBackend ([issue#14313](#), [pr#8573](#), Samuel Just)
- rbd: snap rollback: restore the link to parent ([issue#14512](#), [pr#8535](#), Alexey Sheplyakov)
- rgw: S3: set EncodingType in ListBucketResult ([issue#15896](#), [pr#8987](#), Victor Makarov, Robin H. Johnson)
- rgw: backport rgwx-copy-if-newer for radosgw-agent ([issue#16262](#), [pr#9671](#), Yehuda Sadeh)
- rgw: bucket listing following object delete is partial ([issue#14826](#), [pr#10555](#), Orit Wasserman)
- rgw: convert plain object to versioned (with null version) when removing ([issue#15243](#), [pr#8755](#), Yehuda Sadeh)
- rgw: fix multi-delete query param parsing. ([issue#16618](#), [pr#10189](#), Robin H. Johnson)
- rgw: have a flavor of bucket deletion to bypass GC and to trigger ([issue#15557](#), [pr#10509](#), Pavan Rallabhandi)
- rgw: keep track of written\_objs correctly ([issue#15886](#), [pr#9240](#), Yehuda Sadeh)

- rgw: multipart ListPartsResult has missing quotes on ETag ([issue#15334](#), [pr#8475](#), xie xingguo, Robin H. Johnson)
- rgw: no Last-Modified, Content-Size and X-Object-Manifest headers if no segments in DLO manifest ([issue#15812](#), [pr#9402](#), Radoslaw Zarzynski)
- rgw: radosgw server abort when user passed bad parameters to set quota ([issue#14190](#), [issue#14191](#), [pr#8313](#), Dunrong Huang)
- rgw: radosgw-admin region-map set is not reporting the bucket quota correctly ([issue#16815](#), [pr#10554](#), Yehuda Sadeh, Orit Wasserman)
- rgw: refrain from sending Content-Type/Content-Length for 304 responses ([issue#16327](#), [issue#13582](#), [issue#15119](#), [issue#14005](#), [pr#8379](#), Yehuda Sadeh, Nathan Cutler, Wido den Hollander)
- rgw: remove bucket index objects when deleting the bucket ([issue#16412](#), [pr#10530](#), Orit Wasserman)
- rgw: set Access-Control-Allow-Origin to an asterisk if allowed in a rule ([issue#15348](#), [pr#8528](#), Wido den Hollander)
- rgw: subset of uploaded objects via radosgw are unretrievable when using EC pool ([issue#15745](#), [pr#9407](#), Yehuda Sadeh)
- rgw: subuser rm fails with status 125 ([issue#14375](#), [pr#9961](#), Orit Wasserman)
- rgw: the swift key remains after removing a subuser ([issue#12890](#), [issue#14375](#), [pr#10718](#), Orit Wasserman, Sangdi Xu)
- rgw: user quota may not adjust on bucket removal ([issue#14507](#), [pr#8113](#), Edward Yang)
- tests: be more generous with test timeout ([issue#15403](#), [pr#8470](#), Loic Dachary)
- tests: qa/workunits/rbd: respect RBD\_CREATE\_ARGS environment variable ([issue#16289](#), [pr#9722](#), Mykola Golub)

## v0.94.7 Hammer

---

This Hammer point release fixes several minor bugs. It also includes a backport of an improved ‘ceph osd reweight-by-utilization’ command for handling OSDs with higher-than-average utilizations.

We recommend that all hammer v0.94.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- auth: keyring permissions for mon deamon ([issue#14950](#), [pr#8049](#), Owen Synge)
- auth: PK11\_DestroyContext() is called twice if PK11\_DigestFinal() fails ([issue#14958](#), [pr#7922](#), Brad Hubbard, Dunrong Huang)
- auth: use libnss more safely ([issue#14620](#), [pr#7488](#), Sage Weil)
- ceph-disk: use blkid instead of sgdisk -i ([issue#14080](#), [issue#14094](#), [pr#7475](#), Ilya Dryomov, Loic Dachary)
- ceph-fuse: fix ceph-fuse writing to stale log file after log rotation ([issue#12350](#), [pr#7110](#), Zhi Zhang)
- ceph init script unconditionally sources /lib/lsb/init-functions ([issue#14402](#), [pr#7797](#), Yan, Zheng)
- ceph.in: Notify user that 'tell' can't be used in interactive mode ([issue#14773](#), [pr#7656](#), David Zafman)
- ceph-objectstore-tool, osd: Fix import handling ([issue#10794](#), [issue#13382](#), [pr#7917](#), Sage Weil, David Zafman)
- client: added permission check based on getgroupelist ([issue#13268](#), [pr#6604](#), Yan, Zheng, Danny Al-Gaaf)
- client: inoderef ([issue#13729](#), [pr#6551](#), Yan, Zheng)
- common: clock skew report is incorrect by ceph health detail command ([issue#14175](#), [pr#8051](#), Joao Eduardo Luis)
- global/pidfile: do not start two daemons with a single pid-file ([issue#13422](#), [pr#7671](#), Loic Dachary, shun song)
- librados: segfault in Objecter::handle\_watch\_notify ([issue#13805](#), [pr#7992](#), Sage Weil)
- librbd: flattening an rbd image with active IO can lead to hang ([issue#14092](#), [issue#14483](#), [pr#7485](#), Jason Dillaman)
- librbd: possible QEMU deadlock after creating image snapshots ([issue#14988](#), [pr#8011](#), Jason Dillaman)
- mon: Bucket owner isn't changed after unlink/link ([issue#11076](#), [pr#8583](#), Zengran Zhang)
- monclient: avoid key renew storm on clock skew ([issue#12065](#), [pr#8398](#), Alexey Sheplyakov)
- mon: implement reweight-by-utilization feature ([issue#15054](#), [pr#8026](#), Kefu Chai, Dan van der Ster, Sage Weil)
- mon/LogMonitor: use the configured facility if log to syslog ([issue#13748](#),

pr#7648, Kefu Chai)

- mon: mon sync does not copy config-key ([issue#14577](#), [pr#7576](#), Xiaowei Chen)
- mon/OSDMonitor: avoid underflow in reweight-by-utilization if max\_change=1 ([issue#15655](#), [pr#8979](#), Samuel Just)
- osd: consume\_maps clearing of waiting\_for\_pg needs to check the spg\_t shard for acting set membership ([issue#14278](#), [pr#7577](#), Samuel Just)
- osd: log inconsistent shard sizes ([issue#14009](#), [pr#6946](#), Loic Dachary)
- osd: OSD coredumps with leveldb compact on mount = true ([issue#14748](#), [pr#7645](#), Xiaoxi Chen)
- osd/OSDMap: reset osd\_primary\_affinity shared\_ptr when deepish\_copy\_from ([issue#14686](#), [pr#7590](#), Xinze Chi)
- osd: Protect against excessively large object map sizes ([issue#15121](#), [pr#8401](#), Jason Dillaman)
- osd/ReplicatedPG: do not proxy read *and* process op locally ([issue#15171](#), [pr#8187](#), Sage Weil)
- osd: scrub bogus results when missing a clone ([issue#14875](#), [issue#14874](#), [issue#14877](#), [issue#10098](#), [issue#14878](#), [issue#14881](#), [issue#14882](#), [issue#14883](#), [issue#14879](#), [issue#10290](#), [issue#12740](#), [issue#12738](#), [issue#14880](#), [issue#11135](#), [issue#14876](#), [issue#10809](#), [issue#12193](#), [issue#11237](#), [pr#7702](#), Xinze Chi, Sage Weil, John Spray, Kefu Chai, Mykola Golub, David Zafman)
- osd: Unable to bring up OSD's after dealing with FULL cluster (OSD assert with /include/interval\_set.h: 386: FAILED assert(\_size >= 0)) ([issue#14428](#), [pr#7415](#), Alexey Sheplyakov)
- osd: use GMT time for the object name of hitsets ([issue#13192](#), [issue#9732](#), [issue#12968](#), [pr#7883](#), Kefu Chai, David Zafman)
- qa/workunits/post-file.sh: sudo ([issue#14586](#), [pr#7456](#), Sage Weil)
- qa/workunits: remove 'mds setmap' from workunits ([pr#8123](#), Sage Weil)
- rgw: default quota params ([issue#12997](#), [pr#7188](#), Daniel Gryniewicz)
- rgw: make rgw\_fronends more forgiving of whitespace ([issue#12038](#), [pr#7414](#), Matt Benjamin)
- rgw: radosgw-admin bucket check -fix not work ([issue#14215](#), [pr#7185](#), Weijun Duan)
- rpm package building fails if the build machine has lttng and babeltrace development packages installed locally ([issue#14844](#), [pr#8440](#), Kefu Chai)
- rpm: redhat-lsb-core dependency was dropped, but is still needed ([issue#14906](#),

- pr#7876, Nathan Cutler)
- test\_bit\_vector.cc uses magic numbers against #defines that vary (issue#14747, pr#7672, Jason Dillaman)
- test/librados/tier.cc doesn't completely clean up EC pools (issue#13878, pr#8052, Loic Dachary, Dan Mick)
- tests: bufferlist: do not expect !is\_page\_aligned() after unaligned rebuild (issue#15305, pr#8272, Kefu Chai)
- tools: fix race condition in seq/rand bench (part 1) (issue#14968, issue#14873, pr#7896, Alexey Sheplyakov, Piotr Dałek)
- tools: fix race condition in seq/rand bench (part 2) (issue#14873, pr#7817, Alexey Sheplyakov)
- tools/rados: add bench smoke tests (issue#14971, pr#7903, Piotr Dałek)
- tools, test: Add ceph-objectstore-tool to operate on the meta collection (issue#14977, pr#7911, David Zafman)
- unittest\_crypto: benchmark 100,000 CryptoKey::encrypt() calls (issue#14863, pr#7801, Sage Weil)

## v0.94.6 Hammer

---

This Hammer point release fixes a range of bugs, most notably a fix for unbounded growth of the monitor's leveldb store, and a workaround in the OSD to keep most xattrs small enough to be stored inline in XFS inodes.

We recommend that all hammer v0.94.x users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build/ops: Ceph daemon failed to start, because the service name was already used. (issue#13474, pr#6832, Chuanhong Wang)
- build/ops: LTTng-UST tracing should be dynamically enabled (issue#13274, pr#6415, Jason Dillaman)
- build/ops: ceph upstart script rbdmap.conf incorrectly processes parameters (issue#13214, pr#6159, Sage Weil)
- build/ops: ceph.spec.in License line does not reflect COPYING (issue#12935, pr#6680, Nathan Cutler)

- build/ops: ceph.spec.in libcephfs\_jni1 has no %post and %postun ([issue#12927](#), [pr#5789](#), Owen Synge)
- build/ops: configure.ac: no use to add "+" before ac\_ext=c ([issue#14330](#), [pr#6973](#), Kefu Chai, Robin H. Johnson)
- build/ops: deb: strip tracepoint libraries from Wheezy/Precise builds ([issue#14801](#), [pr#7316](#), Jason Dillaman)
- build/ops: init script reload doesn't work on EL7 ([issue#13709](#), [pr#7187](#), Hervé Rousseau)
- build/ops: init-rbdmap uses distro-specific functions ([issue#12415](#), [pr#6528](#), Boris Ranto)
- build/ops: logrotate reload error on Ubuntu 14.04 ([issue#11330](#), [pr#5787](#), Sage Weil)
- build/ops: miscellaneous spec file fixes ([issue#12931](#), [issue#12994](#), [issue#12924](#), [issue#12360](#), [pr#5790](#), Boris Ranto, Nathan Cutler, Owen Synge, Travis Rhoden, Ken Dreyer)
- build/ops: pass tcmalloc env through to ceph-os ([issue#14802](#), [pr#7365](#), Sage Weil)
- build/ops: rbd-replay-\* moved from ceph-test-dbg to ceph-common-dbg as well ([issue#13785](#), [pr#6580](#), Loic Dachary)
- build/ops: unknown argument -quiet in udevadm settle ([issue#13560](#), [pr#6530](#), Jason Dillaman)
- common: Objecter: pool op callback may hang forever. ([issue#13642](#), [pr#6588](#), xie xingguo)
- common: Objecter: potential null pointer access when do pool\_snap\_list. ([issue#13639](#), [pr#6839](#), xie xingguo)
- common: ThreadPool add/remove work queue methods not thread safe ([issue#12662](#), [pr#5889](#), Jason Dillaman)
- common: auth/cephx: large amounts of log are produced by osd ([issue#13610](#), [pr#6835](#), Qiankun Zheng)
- common: client nonce collision due to unshared pid namespaces ([issue#13032](#), [pr#6151](#), Josh Durgin)
- common: common/Thread:pthread\_attr\_destroy(thread\_attr) when done with it ([issue#12570](#), [pr#6157](#), Piotr Dałek)
- common: log: Log.cc: Assign LOG\_DEBUG priority to syslog calls ([issue#13993](#), [pr#6994](#), Brad Hubbard)

- common: objecter: cancellation bugs ([issue#13071](#), [pr#6155](#), Jianpeng Ma)
- common: pure virtual method called ([issue#13636](#), [pr#6587](#), Jason Dillaman)
- common: small probability sigabrt when setting rados\_osd\_op\_timeout ([issue#13208](#), [pr#6143](#), Ruifeng Yang)
- common: wrong conditional for boolean function KeyServer::get\_auth() ([issue#9756](#), [issue#13424](#), [pr#6213](#), Nathan Cutler)
- crush: crash if we see CRUSH\_ITEM\_NONE in early rule step ([issue#13477](#), [pr#6430](#), Sage Weil)
- doc: man: document listwatchers cmd in “rados” manpage ([issue#14556](#), [pr#7434](#), Kefu Chai)
- doc: regenerate man pages, add orphans commands to radosgw-admin(8) ([issue#14637](#), [pr#7524](#), Ken Dreyer)
- fs: CephFS restriction on removing cache tiers is overly strict ([issue#11504](#), [pr#6402](#), John Spray)
- fs: fsstress.sh fails ([issue#12710](#), [pr#7454](#), Yan, Zheng)
- librados: LibRadosWatchNotify.WatchNotify2Timeout ([issue#13114](#), [pr#6336](#), Sage Weil)
- librbd: ImageWatcher shouldn’t block the notification thread ([issue#14373](#), [pr#7407](#), Jason Dillaman)
- librbd: diff\_iterate needs to handle holes in parent images ([issue#12885](#), [pr#6097](#), Jason Dillaman)
- librbd: fix merge-diff for >2GB diff-files ([issue#14030](#), [pr#6980](#), Jason Dillaman)
- librbd: invalidate object map on error even w/o holding lock ([issue#13372](#), [pr#6289](#), Jason Dillaman)
- librbd: reads larger than cache size hang ([issue#13164](#), [pr#6354](#), Lu Shi)
- mds: ceph mds add\_data\_pool check for EC pool is wrong ([issue#12426](#), [pr#5766](#), John Spray)
- mon: MonitorDBStore: get\_next\_key() only if prefix matches ([issue#11786](#), [pr#5361](#), Joao Eduardo Luis)
- mon: OSDMonitor: do not assume a session exists in send\_incremental() ([issue#14236](#), [pr#7150](#), Joao Eduardo Luis)
- mon: check for store writeability before participating in election ([issue#13089](#), [pr#6144](#), Sage Weil)

- mon: compact full epochs also ([issue#14537](#), [pr#7446](#), Kefu Chai)
- mon: include min\_last\_epoch\_clean as part of PGMap::print\_summary and PGMap::dump ([issue#13198](#), [pr#6152](#), Guang Yang)
- mon: map\_cache can become inaccurate if osd does not receive the osdmaps ([issue#10930](#), [pr#5773](#), Kefu Chai)
- mon: should not set isvalid = true when cephx\_verify\_authorizer return false ([issue#13525](#), [pr#6391](#), Ruifeng Yang)
- osd: Ceph Pools' MAX AVAIL is 0 if some OSDs' weight is 0 ([issue#13840](#), [pr#6834](#), Chengyuan Li)
- osd: FileStore calls syncfs(2) even it is not supported ([issue#12512](#), [pr#5530](#), Kefu Chai)
- osd: FileStore: potential memory leak if getattrs fails. ([issue#13597](#), [pr#6420](#), xie xingguo)
- osd: IO error on kvm/rbd with an erasure coded pool tier ([issue#12012](#), [pr#5897](#), Kefu Chai)
- osd: OSD::build\_past\_intervals\_parallel() shall reset primary and up\_primary when begin a new past\_interval. ([issue#13471](#), [pr#6398](#), xiexingguo)
- osd: ReplicatedBackend: populate recovery\_info.size for clone (bug symptom is size mismatch on replicated backend on a clone in scrub) ([issue#12828](#), [pr#6153](#), Samuel Just)
- osd: ReplicatedPG: wrong result code checking logic during sparse\_read ([issue#14151](#), [pr#7179](#), xie xingguo)
- osd: ReplicatedPG::hit\_set\_trim osd/ReplicatedPG.cc: 11006: FAILED assert(obc) ([issue#13192](#), [issue#9732](#), [issue#12968](#), [pr#5825](#), Kefu Chai, Zhiqiang Wang, Samuel Just, David Zafman)
- osd: avoid multi set osd\_op.outdata in tier pool ([issue#12540](#), [pr#6060](#), Xinze Chi)
- osd: bug with cache/tiering and snapshot reads ([issue#12748](#), [pr#6589](#), Kefu Chai)
- osd: ceph osd pool stats broken in hammer ([issue#13843](#), [pr#7180](#), BJ Lougee)
- osd: ceph-disk prepare fails if device is a symlink ([issue#13438](#), [pr#7176](#), Joe Julian)
- osd: check for full before changing the cached obc (hammer) ([issue#13098](#), [pr#6918](#), Alexey Sheplyakov)
- osd: config\_opts: increase suicide timeout to 300 to match recovery ([issue#14376](#),

pr#7236, Samuel Just)

- osd: disable filestore\_xfs\_extsize by default ([issue#14397](#), [pr#7411](#), Ken Dreyer)
- osd: do not cache unused memory in attrs ([issue#12565](#), [pr#6499](#), Xinze Chi, Ning Yao)
- osd: dumpling incrementals do not work properly on hammer and newer ([issue#13234](#), [pr#6132](#), Samuel Just)
- osd: filestore: fix peek\_queue for OpSequencer ([issue#13209](#), [pr#6145](#), Xinze Chi)
- osd: hit set clear repops fired in same epoch as map change - segfault since they fall into the new interval even though the repops are cleared ([issue#12809](#), [pr#5890](#), Samuel Just)
- osd: object\_info\_t::decode() has wrong version ([issue#13462](#), [pr#6335](#), David Zafman)
- osd: osd/OSD.cc: 2469: FAILED assert(pg\_stat\_queue.empty()) on shutdown ([issue#14212](#), [pr#7178](#), Sage Weil)
- osd: osd/PG.cc: 288: FAILED assert(info.last\_epoch\_started >= info.history.last\_epoch\_started) ([issue#14015](#), [pr#7177](#), David Zafman)
- osd: osd/PG.cc: 3837: FAILED assert(0 == "Running incompatible OSD") ([issue#11661](#), [pr#7206](#), David Zafman)
- osd: osd/ReplicatedPG: Recency fix ([issue#14320](#), [pr#7207](#), Sage Weil, Robert LeBlanc)
- osd: pg stuck in replay ([issue#13116](#), [pr#6401](#), Sage Weil)
- osd: race condition detected during send\_failures ([issue#13821](#), [pr#6755](#), Sage Weil)
- osd: randomize scrub times ([issue#10973](#), [pr#6199](#), Kefu Chai)
- osd: requeue\_scrub when kick\_object\_context\_blocked ([issue#12515](#), [pr#5891](#), Xinze Chi)
- osd: revert: use GMT time for hitsets ([issue#13812](#), [pr#6644](#), Loic Dachary)
- osd: segfault in agent\_work ([issue#13199](#), [pr#6146](#), Samuel Just)
- osd: should recalc the min\_last\_epoch\_clean when decode PGMap ([issue#13112](#), [pr#6154](#), Kefu Chai)
- osd: smaller object\_info\_t xattrs ([issue#14803](#), [pr#6544](#), Sage Weil)
- osd: we do not ignore notify from down osds ([issue#12990](#), [pr#6158](#), Samuel Just)

- rbd: QEMU hangs after creating snapshot and stopping VM ([issue#13726](#), [pr#6586](#), Jason Dillaman)
- rbd: TaskFinisher::cancel should remove event from SafeTimer ([issue#14476](#), [pr#7417](#), Douglas Fuller)
- rbd: avoid re-writing old-format image header on resize ([issue#13674](#), [pr#6585](#), Jason Dillaman)
- rbd: fix bench-write ([issue#14225](#), [pr#7183](#), Sage Weil)
- rbd: rbd-replay does not check for EOF and goes to endless loop ([issue#14452](#), [pr#7416](#), Mykola Golub)
- rbd: rbd-replay-prep and rbd-replay improvements ([issue#13221](#), [issue#13220](#), [issue#13378](#), [pr#6286](#), Jason Dillaman)
- rbd: verify self-managed snapshot functionality on image create ([issue#13633](#), [pr#7182](#), Jason Dillaman)
- rgw: Make RGW\_MAX\_PUT\_SIZE configurable ([issue#6999](#), [pr#7441](#), Vladislav Odintsov, Yuan Zhou)
- rgw: Setting ACL on Object removes ETag ([issue#12955](#), [pr#6620](#), Brian Felton)
- rgw: backport content-type casing ([issue#12939](#), [pr#5910](#), Robin H. Johnson)
- rgw: bucket listing hangs on versioned buckets ([issue#12913](#), [pr#6352](#), Yehuda Sadeh)
- rgw: fix wrong etag calculation during POST on S3 bucket. ([issue#11241](#), [pr#7442](#), Vladislav Odintsov, Radoslaw Zarzynski)
- rgw: get bucket location returns region name, not region api name ([issue#13458](#), [pr#6349](#), Yehuda Sadeh)
- rgw: missing handling of encoding-type=url when listing keys in bucket ([issue#12735](#), [pr#6527](#), Jeff Weber)
- rgw: orphan tool should be careful about removing head objects ([issue#12958](#), [pr#6351](#), Yehuda Sadeh)
- rgw: orphans finish segfaults ([issue#13824](#), [pr#7186](#), Igor Fedotov)
- rgw: rgw-admin: document orphans commands in usage ([issue#14516](#), [pr#7526](#), Yehuda Sadeh)
- rgw: swift API returns more than real object count and bytes used when retrieving account metadata ([issue#13140](#), [pr#6512](#), Sangdi Xu)
- rgw: swift use Civetweb ssl can not get right url ([issue#13628](#), [pr#6491](#), Weijun Duan)

- rgw: value of Swift API's X-Object-Manifest header is not url\_decoded during segment look up ([issue#12728](#), [pr#6353](#), Radoslaw Zarzynski)
- tests: fixed broken Makefiles after integration of ttng into rados ([issue#13210](#), [pr#6322](#), Sebastien Ponce)
- tests: fsx failed to compile ([issue#14384](#), [pr#7501](#), Greg Farnum)
- tests: notification slave needs to wait for master ([issue#13810](#), [pr#7226](#), Jason Dillaman)
- tests: qa: remove legacy OS support from rbd/qemu-iotests ([issue#13483](#), [issue#14385](#), [pr#7252](#), Vasu Kulkarni, Jason Dillaman)
- tests: testprofile must be removed before it is re-created ([issue#13664](#), [pr#6450](#), Loic Dachary)
- tools: ceph-monstore-tool must do out\_store.close() ([issue#10093](#), [pr#7347](#), huangjun)
- tools: heavy memory shuffling in rados bench ([issue#12946](#), [pr#5810](#), Piotr Dałek)
- tools: race condition in rados bench ([issue#12947](#), [pr#6791](#), Piotr Dałek)
- tools: tool for artificially inflate the leveldb of the mon store for testing purposes ([issue#10093](#), [issue#11815](#), [issue#14217](#), [pr#7412](#), Cilang Zhao, Bo Cai, Kefu Chai, huangjun, Joao Eduardo Luis)

## v0.94.5 Hammer

---

This Hammer point release fixes a critical regression in librbd that can cause QEMU/KVM to crash when caching is enabled on images that have been cloned.

All v0.94.4 Hammer users are strongly encouraged to upgrade.

## Notable Changes

---

- librbd: potential assertion failure during cache read ([issue#13559](#), [pr#6348](#), Jason Dillaman)
- osd: osd/ReplicatedPG: remove stray debug line ([issue#13455](#), [pr#6362](#), Sage Weil)
- tests: qemu workunit refers to apt-mirror.front.sepia.ceph.com ([issue#13420](#), [pr#6330](#), Yuan Zhou)

For more detailed information, see [the complete changelog](#).

## v0.94.4 Hammer

---

This Hammer point release fixes several important bugs in Hammer, as well as fixing interoperability issues that are required before an upgrade to Infernalis. That is, all users of earlier version of Hammer or any version of Firefly will first need to upgrade to hammer v0.94.4 or later before upgrading to Infernalis (or future releases).

All v0.94.x Hammer users are strongly encouraged to upgrade.

## Notable Changes

---

- build/ops: ceph.spec.in: 50-rbd.rules conditional is wrong ([issue#12166](#), [pr#5207](#), Nathan Cutler)
- build/ops: ceph.spec.in: ceph-common needs python-argparse on older distros, but doesn't require it ([issue#12034](#), [pr#5216](#), Nathan Cutler)
- build/ops: ceph.spec.in: radosgw requires apache for SUSE only - makes no sense ([issue#12358](#), [pr#5411](#), Nathan Cutler)
- build/ops: ceph.spec.in: rpm: cephfs\_java not fully conditionalized ([issue#11991](#), [pr#5202](#), Nathan Cutler)
- build/ops: ceph.spec.in: rpm: not possible to turn off Java ([issue#11992](#), [pr#5203](#), Owen Synge)
- build/ops: ceph.spec.in: running fdupes unnecessarily ([issue#12301](#), [pr#5223](#), Nathan Cutler)
- build/ops: ceph.spec.in: snappy-devel for all supported distros ([issue#12361](#), [pr#5264](#), Nathan Cutler)
- build/ops: ceph.spec.in: SUSE/openSUSE builds need libbz2-devel ([issue#11629](#), [pr#5204](#), Nathan Cutler)
- build/ops: ceph.spec.in: useless %py\_requires breaks SLE11-SP3 build ([issue#12351](#), [pr#5412](#), Nathan Cutler)
- build/ops: error in ext\_mime\_map\_init() when /etc/mime.types is missing ([issue#11864](#), [pr#5385](#), Ken Dreyer)
- build/ops: upstart: limit respawn to 3 in 30 mins (instead of 5 in 30s) ([issue#11798](#), [pr#5930](#), Sage Weil)
- build/ops: With root as default user, unable to have multiple RGW instances running ([issue#10927](#), [pr#6161](#), Sage Weil)
- build/ops: With root as default user, unable to have multiple RGW instances running ([issue#11140](#), [pr#6161](#), Sage Weil)
- build/ops: With root as default user, unable to have multiple RGW instances

- running ([issue#11686](#), [pr#6161](#), Sage Weil)
- build/ops: With root as default user, unable to have multiple RGW instances running ([issue#12407](#), [pr#6161](#), Sage Weil)
- cli: ceph: cli throws exception on unrecognized errno ([issue#11354](#), [pr#5368](#), Kefu Chai)
- cli: ceph tell: broken error message / misleading hinting ([issue#11101](#), [pr#5371](#), Kefu Chai)
- common: arm: all programs that link to librados2 hang forever on startup ([issue#12505](#), [pr#5366](#), Boris Ranto)
- common: buffer: critical bufferlist::zero bug ([issue#12252](#), [pr#5365](#), Haomai Wang)
- common: ceph-object-corpus: add 0.94.2-207-g88e7ee7 hammer objects ([issue#13070](#), [pr#5551](#), Sage Weil)
- common: do not insert empty ptr when rebuild empty bufferlist ([issue#12775](#), [pr#5764](#), Xinze Chi)
- common: [ FAILED ] TestLibRBD.BlockingAIO ([issue#12479](#), [pr#5768](#), Jason Dillaman)
- common: LibCephFS.GetPoolId failure ([issue#12598](#), [pr#5887](#), Yan, Zheng)
- common: Memory leak in Mutex.cc, pthread\_mutexattr\_init without pthread\_mutexattr\_destroy ([issue#11762](#), [pr#5378](#), Ketor Meng)
- common: object\_map\_update fails with -EINVAL return code ([issue#12611](#), [pr#5559](#), Jason Dillaman)
- common: Pipe: Drop connect\_seq increase line ([issue#13093](#), [pr#5908](#), Haomai Wang)
- common: recursive lock of md\_config\_t (0) ([issue#12614](#), [pr#5759](#), Josh Durgin)
- crush: ceph osd crush reweight-subtree does not reweight parent node ([issue#11855](#), [pr#5374](#), Sage Weil)
- doc: update docs to point to download.ceph.com ([issue#13162](#), [pr#6156](#), Alfredo Deza)
- fs: ceph-fuse 0.94.2-1trusty segfaults / aborts ([issue#12297](#), [pr#5381](#), Greg Farnum)
- fs: segfault launching ceph-fuse with bad -name ([issue#12417](#), [pr#5382](#), John Spray)
- librados: Change radosgw pools default crush ruleset ([issue#11640](#), [pr#5754](#), Yuan Zhou)
- librbd: correct issues discovered via lockdep / helgrind ([issue#12345](#), [pr#5296](#),

Jason Dillaman)

- librbd: Crash during TestInternal.MultipleResize ([issue#12664](#), [pr#5769](#), Jason Dillaman)
- librbd: deadlock during cooperative exclusive lock transition ([issue#11537](#), [pr#5319](#), Jason Dillaman)
- librbd: Possible crash while concurrently writing and shrinking an image ([issue#11743](#), [pr#5318](#), Jason Dillaman)
- mon: add a cache layer over MonitorDBStore ([issue#12638](#), [pr#5697](#), Kefu Chai)
- mon: fix crush testing for new pools ([issue#13400](#), [pr#6192](#), Sage Weil)
- mon: get pools health'info have error ([issue#12402](#), [pr#5369](#), renhwztetecs)
- mon: implicit erasure code crush ruleset is not validated ([issue#11814](#), [pr#5276](#), Loic Dachary)
- mon: PaxosService: call post\_refresh() instead of post\_paxos\_update() ([issue#11470](#), [pr#5359](#), Joao Eduardo Luis)
- mon: pgmonitor: wrong at/near target max" reporting ([issue#12401](#), [pr#5370](#), huangjun)
- mon: register\_new\_pgs() should check ruleno instead of its index ([issue#12210](#), [pr#5377](#), Xinze Chi)
- mon: Show osd as NONE in ceph osd map <pool> <object> output ([issue#11820](#), [pr#5376](#), Shylesh Kumar)
- mon: the output is wrong when runing ceph osd reweight ([issue#12251](#), [pr#5372](#), Joao Eduardo Luis)
- osd: allow peek\_map\_epoch to return an error ([issue#13060](#), [pr#5892](#), Sage Weil)
- osd: cache agent is idle although one object is left in the cache ([issue#12673](#), [pr#5765](#), Loic Dachary)
- osd: copy-from doesn't preserve truncate\_{seq,size} ([issue#12551](#), [pr#5885](#), Samuel Just)
- osd: crash creating/deleting pools ([issue#12429](#), [pr#5527](#), John Spray)
- osd: fix repair when recorded digest is wrong ([issue#12577](#), [pr#5468](#), Sage Weil)
- osd: include/ceph\_features: define HAMMER\_0\_94\_4 feature ([issue#13026](#), [pr#5687](#), Sage Weil)
- osd: is\_new\_interval() fixes ([issue#10399](#), [pr#5691](#), Jason Dillaman)

- osd: is\_new\_interval() fixes ([issue#11771](#), [pr#5691](#), Jason Dillaman)
- osd: long standing slow requests: connection->session->waiting\_for\_map->connection ref cycle ([issue#12338](#), [pr#5761](#), Samuel Just)
- osd: Mutex Assert from PipeConnection::try\_get\_pipe ([issue#12437](#), [pr#5758](#), David Zafman)
- osd: pg\_interval\_t::check\_new\_interval - for ec pool, should not rely on min\_size to determine if the PG was active at the interval ([issue#12162](#), [pr#5373](#), Guang G Yang)
- osd: PGLog.cc: 732: FAILED assert(log.log.size() == log\_keys\_debug.size()) ([issue#12652](#), [pr#5763](#), Sage Weil)
- osd: PGLog::proc\_replica\_log: correctly handle case where entries between olog.head and log.tail were split out ([issue#11358](#), [pr#5380](#), Samuel Just)
- osd: read on chunk-aligned xattr not handled ([issue#12309](#), [pr#5367](#), Sage Weil)
- osd: suicide timeout during peering - search for missing objects ([issue#12523](#), [pr#5762](#), Guang G Yang)
- osd: WBThrottle::clear\_object: signal on cond when we reduce throttle values ([issue#12223](#), [pr#5757](#), Samuel Just)
- rbd: crash during shutdown after writeback blocked by IO errors ([issue#12597](#), [pr#5767](#), Jianpeng Ma)
- rgw: add delimiter to prefix only when path is specified ([issue#12960](#), [pr#5860](#), Sylvain Baubéau)
- rgw: create a tool for orphaned objects cleanup ([issue#9604](#), [pr#5717](#), Yehuda Sadeh)
- rgw: don't preserve acls when copying object ([issue#11563](#), [pr#6039](#), Yehuda Sadeh)
- rgw: don't preserve acls when copying object ([issue#12370](#), [pr#6039](#), Yehuda Sadeh)
- rgw: don't preserve acls when copying object ([issue#13015](#), [pr#6039](#), Yehuda Sadeh)
- rgw: Ensure that swift keys don't include backslashes ([issue#7647](#), [pr#5716](#), Yehuda Sadeh)
- rgw: GWWatcher::handle\_error -> common/Mutex.cc: 95: FAILED assert(r == 0) ([issue#12208](#), [pr#6164](#), Yehuda Sadeh)
- rgw: HTTP return code is not being logged by CivetWeb ([issue#12432](#), [pr#5498](#), Yehuda Sadeh)
- rgw: init\_rados failed leads to repeated delete ([issue#12978](#), [pr#6165](#), Xiaowei Chen)

- rgw: init some manifest fields when handling explicit objs ([issue#11455](#), [pr#5732](#), Yehuda Sadeh)
- rgw: Keystone Fernet tokens break auth ([issue#12761](#), [pr#6162](#), Abhishek Lekshmanan)
- rgw: region data still exist in region-map after region-map update ([issue#12964](#), [pr#6163](#), dwj192)
- rgw: remove trailing :port from host for purposes of subdomain matching ([issue#12353](#), [pr#6042](#), Yehuda Sadeh)
- rgw: rest-bench common/WorkQueue.cc: 54: FAILED assert(\_threads.empty()) ([issue#3896](#), [pr#5383](#), huangjun)
- rgw: returns requested bucket name raw in Bucket response header ([issue#12537](#), [pr#5715](#), Yehuda Sadeh)
- rgw: segmentation fault when rgw\_gc\_max\_objs > HASH\_PRIME ([issue#12630](#), [pr#5719](#), Ruifeng Yang)
- rgw: segments are read during HEAD on Swift DLO ([issue#12780](#), [pr#6160](#), Yehuda Sadeh)
- rgw: setting max number of buckets for user via ceph.conf option ([issue#12714](#), [pr#6166](#), Vikhyat Umrao)
- rgw: Swift API: X-Trans-Id header is wrongly formatted ([issue#12108](#), [pr#5721](#), Radoslaw Zarzynski)
- rgw: testGetContentType and testHead failed ([issue#11091](#), [pr#5718](#), Radoslaw Zarzynski)
- rgw: testGetContentType and testHead failed ([issue#11438](#), [pr#5718](#), Radoslaw Zarzynski)
- rgw: testGetContentType and testHead failed ([issue#12157](#), [pr#5718](#), Radoslaw Zarzynski)
- rgw: testGetContentType and testHead failed ([issue#12158](#), [pr#5718](#), Radoslaw Zarzynski)
- rgw: testGetContentType and testHead failed ([issue#12363](#), [pr#5718](#), Radoslaw Zarzynski)
- rgw: the arguments 'domain' should not be assigned when return false ([issue#12629](#), [pr#5720](#), Ruifeng Yang)
- tests: qa/workunits/cephtool/test.sh: don't assume crash\_replay\_interval=45 ([issue#13406](#), [pr#6172](#), Sage Weil)

- tests: TEST\_crush\_rule\_create\_erasure consistently fails on i386 builder ([issue#12419](#), [pr#6201](#), Loic Dachary)
- tools: ceph-disk zap should ensure block device ([issue#11272](#), [pr#5755](#), Loic Dachary)

For more detailed information, see [the complete changelog](#).

## v0.94.3 Hammer

---

This Hammer point release fixes a critical (though rare) data corruption bug that could be triggered when logs are rotated via SIGHUP. It also fixes a range of other important bugs in the OSD, monitor, RGW, RGW, and CephFS.

All v0.94.x Hammer users are strongly encouraged to upgrade.

## Upgrading

---

- The `pg ls-by-{pool,primary,osd}` commands and `pg ls` now take the argument `recovering` instead of `recovery` in order to include the recovering pgs in the listed pgs.

## Notable Changes

---

- librbd: aio calls may block ([issue#11770](#), [pr#4875](#), Jason Dillaman)
- osd: make the all osd/filestore thread pool suicide timeouts separately configurable ([issue#11701](#), [pr#5159](#), Samuel Just)
- mon: ceph fails to compile with boost 1.58 ([issue#11982](#), [pr#5122](#), Kefu Chai)
- tests: TEST\_crush\_reject\_empty must not run a mon ([issue#12285](#), [issue#11975](#), [pr#5208](#), Kefu Chai)
- osd: FAILED assert(!old\_value.deleted()) in upgrade:giant-x-hammer-distro-basic-multi run ([issue#11983](#), [pr#5121](#), Samuel Just)
- build/ops: linking ceph to tcmalloc causes segfault on SUSE SLE11-SP3 ([issue#12368](#), [pr#5265](#), Thorsten Behrens)
- common: utf8 and old gcc breakage on RHEL6.5 ([issue#7387](#), [pr#4687](#), Kefu Chai)
- crush: take crashes due to invalid arg ([issue#11740](#), [pr#4891](#), Sage Weil)
- rgw: need conversion tool to handle fixes following #11974 ([issue#12502](#), [pr#5384](#), Yehuda Sadeh)
- rgw: Swift API: support for 202 Accepted response code on container creation ([issue#12299](#), [pr#5214](#), Radoslaw Zarzynski)

- common: Log::reopen\_log\_file: take m\_flush\_mutex ([issue#12520](#), [pr#5405](#), Samuel Just)
- rgw: Properly respond to the Connection header with Civetweb ([issue#12398](#), [pr#5284](#), Wido den Hollander)
- rgw: multipart list part response returns incorrect field ([issue#12399](#), [pr#5285](#), Henry Chang)
- build/ops: ceph.spec.in: 95-ceph-osd.rules, mount.ceph, and mount.fuse.ceph not installed properly on SUSE ([issue#12397](#), [pr#5283](#), Nathan Cutler)
- rgw: radosgw-admin dumps user info twice ([issue#12400](#), [pr#5286](#), guce)
- doc: fix doc build ([issue#12180](#), [pr#5095](#), Kefu Chai)
- tests: backport 11493 fixes, and test, preventing ec cache pools ([issue#12314](#), [pr#4961](#), Samuel Just)
- rgw: does not send Date HTTP header when civetweb frontend is used ([issue#11872](#), [pr#5228](#), Radoslaw Zarzynski)
- mon: pg ls is broken ([issue#11910](#), [pr#5160](#), Kefu Chai)
- librbd: A client opening an image mid-resize can result in the object map being invalidated ([issue#12237](#), [pr#5279](#), Jason Dillaman)
- doc: missing man pages for ceph-create-keys, ceph-disk-\* ([issue#11862](#), [pr#4846](#), Nathan Cutler)
- tools: ceph-post-file fails on rhel7 ([issue#11876](#), [pr#5038](#), Sage Weil)
- build/ops: rcceph script is buggy ([issue#12090](#), [pr#5028](#), Owen Synge)
- rgw: Bucket header is enclosed by quotes ([issue#11874](#), [pr#4862](#), Wido den Hollander)
- build/ops: packaging: add SuSEfirewall2 service files ([issue#12092](#), [pr#5030](#), Tim Serong)
- rgw: Keystone PKI token expiration is not enforced ([issue#11722](#), [pr#4884](#), Anton Aksola)
- build/ops: debian/control: ceph-common (>> 0.94.2) must be >= 0.94.2-2 ([issue#12529](#), [11998](#), [pr#5417](#), Loic Dachary)
- mon: Clock skew causes missing summary and confuses Calamari ([issue#11879](#), [pr#4868](#), Thorsten Behrens)
- rgw: rados objects wronly deleted ([issue#12099](#), [pr#5117](#), wuxingyi)
- tests: kernel\_untar\_build fails on EL7 ([issue#12098](#), [pr#5119](#), Greg Farnum)

- fs: Fh ref count will leak if readahead does not need to do read from osd ([issue#12319](#), [pr#5427](#), Zhi Zhang)
- mon: OSDMonitor: allow addition of cache pool with non-empty snaps with co... ([issue#12595](#), [pr#5252](#), Samuel Just)
- mon: MDSMonitor: handle MDSBeacon messages properly ([issue#11979](#), [pr#5123](#), Kefu Chai)
- tools: ceph-disk: get\_partition\_type fails on /dev/cciss... ([issue#11760](#), [pr#4892](#), islepnev)
- build/ops: max files open limit for OSD daemon is too low ([issue#12087](#), [pr#5026](#), Owen Synge)
- mon: add an “osd crush tree” command ([issue#11833](#), [pr#5248](#), Kefu Chai)
- mon: mon crashes when “ceph osd tree 85 -format json” ([issue#11975](#), [pr#4936](#), Kefu Chai)
- build/ops: ceph / ceph-dbg steal ceph-objecstore-tool from ceph-test / ceph-test-dbg ([issue#11806](#), [pr#5069](#), Loic Dachary)
- rgw: DragonDisk fails to create directories via S3: MissingContentLength ([issue#12042](#), [pr#5118](#), Yehuda Sadeh)
- build/ops: /usr/bin/ceph from ceph-common is broken without installing ceph ([issue#11998](#), [pr#5206](#), Ken Dreyer)
- build/ops: systemd: Increase max files open limit for OSD daemon ([issue#11964](#), [pr#5040](#), Owen Synge)
- build/ops: rgw/logrotate.conf calls service with wrong init script name ([issue#12044](#), [pr#5055](#), wuxingyi)
- common: OPT\_INT option interprets 3221225472 as -1073741824, and crashes in Throttle::Throttle() ([issue#11738](#), [pr#4889](#), Kefu Chai)
- doc: doc/release-notes: v0.94.2 ([issue#11492](#), [pr#4934](#), Sage Weil)
- common: admin\_socket: close socket descriptor in destructor ([issue#11706](#), [pr#4657](#), Jon Bernard)
- rgw: Object copy bug ([issue#11755](#), [pr#4885](#), Javier M. Mellid)
- rgw: empty json response when getting user quota ([issue#12245](#), [pr#5237](#), wuxingyi)
- fs: cephfs Dumper tries to load whole journal into memory at once ([issue#11999](#), [pr#5120](#), John Spray)
- rgw: Fix tool for #11442 does not correctly fix objects created via multipart uploads ([issue#12242](#), [pr#5229](#), Yehuda Sadeh)

- rgw: Civetweb RGW appears to report full size of object as downloaded when only partially downloaded ([issue#12243](#), [pr#5231](#), Yehuda Sadeh)
- osd: stuck incomplete ([issue#12362](#), [pr#5269](#), Samuel Just)
- osd: start\_flush: filter out removed snaps before determining snapc's ([issue#11911](#), [pr#4899](#), Samuel Just)
- librbd: internal.cc: 1967: FAILED assert(watchers.size() == 1) ([issue#12239](#), [pr#5243](#), Jason Dillaman)
- librbd: new QA client upgrade tests ([issue#12109](#), [pr#5046](#), Jason Dillaman)
- librbd: [ FAILED ] TestLibRBD.ExclusiveLockTransition ([issue#12238](#), [pr#5241](#), Jason Dillaman)
- rgw: Swift API: XML document generated in response for GET on account does not contain account name ([issue#12323](#), [pr#5227](#), Radoslaw Zarzynski)
- rgw: keystone does not support chunked input ([issue#12322](#), [pr#5226](#), Hervé Rousseau)
- mds: MDS is crashed (mds/CDir.cc: 1391: FAILED assert(!is\_complete())) ([issue#11737](#), [pr#4886](#), Yan, Zheng)
- cli: ceph: cli interactive mode does not understand quotes ([issue#11736](#), [pr#4776](#), Kefu Chai)
- librbd: add valgrind memory checks for unit tests ([issue#12384](#), [pr#5280](#), Zhiqiang Wang)
- build/ops: admin/build-doc: script fails silently under certain circumstances ([issue#11902](#), [pr#4877](#), John Spray)
- osd: Fixes for rados ops with snaps ([issue#11908](#), [pr#4902](#), Samuel Just)
- build/ops: ceph.spec.in: ceph-common subpackage def needs tweaking for SUSE/openSUSE ([issue#12308](#), [pr#4883](#), Nathan Cutler)
- fs: client: reference counting 'struct Fh' ([issue#12088](#), [pr#5222](#), Yan, Zheng)
- build/ops: ceph.spec: update OpenSUSE BuildRequires ([issue#11611](#), [pr#4667](#), Loic Dachary)

For more detailed information, see [the complete changelog](#).

## v0.94.2 Hammer

---

This Hammer point release fixes a few critical bugs in RGW that can prevent objects starting with underscore from behaving properly and that prevent garbage collection of

deleted objects when using the Civetweb standalone mode.

All v0.94.x Hammer users are strongly encouraged to upgrade, and to make note of the repair procedure below if RGW is in use.

## Upgrading from previous Hammer release

Bug #11442 introduced a change that made rgw objects that start with underscore incompatible with previous versions. The fix to that bug reverts to the previous behavior. In order to be able to access objects that start with an underscore and were created in prior Hammer releases, following the upgrade it is required to run (for each affected bucket):

```
1. $ radosgw-admin bucket check --check-head-obj-locator \
   2.           --bucket=<bucket> [--fix]
```

## Notable changes

- build: compilation error: No high-precision counter available (armhf, powerpc...) (#11432, James Page)
- ceph-dencoder links to libtcmalloc, and shouldn't (#10691, Boris Ranto)
- ceph-disk: disk zap sgdisk invocation (#11143, Owen Synge)
- ceph-disk: use a new disk as journal disk,ceph-disk prepare fail (#10983, Loic Dachary)
- ceph-objectstore-tool should be in the ceph server package (#11376, Ken Dreyer)
- librados: can get stuck in redirect loop if osdmap epoch == last\_force\_op\_resend (#11026, Jianpeng Ma)
- librbd: A retransmit of proxied flatten request can result in -EINVAL (Jason Dillaman)
- librbd: ImageWatcher should cancel in-flight ops on watch error (#11363, Jason Dillaman)
- librbd: Objectcacher setting max object counts too low (#7385, Jason Dillaman)
- librbd: Periodic failure of TestLibRBD.DiffIterateStress (#11369, Jason Dillaman)
- librbd: Queued AIO reference counters not properly updated (#11478, Jason Dillaman)
- librbd: deadlock in image refresh (#5488, Jason Dillaman)
- librbd: notification race condition on snap\_create (#11342, Jason Dillaman)

- mds: Hammer uclient checking (#11510, John Spray)
- mds: remove caps from revoking list when caps are voluntarily released (#11482, Yan, Zheng)
- messenger: double clear of pipe in reaper (#11381, Haomai Wang)
- mon: Total size of OSDs is a maginitude less than it is supposed to be. (#11534, Zhe Zhang)
- osd: don't check order in finish\_proxy\_read (#11211, Zhiqiang Wang)
- osd: handle old semi-deleted pgs after upgrade (#11429, Samuel Just)
- osd: object creation by write cannot use an offset on an erasure coded pool (#11507, Jianpeng Ma)
- rgw: Improve rgw HEAD request by avoiding read the body of the first chunk (#11001, Guang Yang)
- rgw: civetweb is hitting a limit (number of threads 1024) (#10243, Yehuda Sadeh)
- rgw: civetweb should use unique request id (#10295, Orit Wasserman)
- rgw: critical fixes for hammer (#11447, #11442, Yehuda Sadeh)
- rgw: fix swift COPY headers (#10662, #10663, #11087, #10645, Radoslaw Zarzynski)
- rgw: improve performance for large object (multiple chunks) GET (#11322, Guang Yang)
- rgw: init-radosgw: run RGW as root (#11453, Ken Dreyer)
- rgw: keystone token cache does not work correctly (#11125, Yehuda Sadeh)
- rgw: make quota/gc thread configurable for starting (#11047, Guang Yang)
- rgw: make swift responses of RGW return last-modified, content-length, x-trans-id headers. (#10650, Radoslaw Zarzynski)
- rgw: merge manifests correctly when there's prefix override (#11622, Yehuda Sadeh)
- rgw: quota not respected in POST object (#11323, Sergey Arkhipov)
- rgw: restore buffer of multipart upload after EEXIST (#11604, Yehuda Sadeh)
- rgw: shouldn't need to disable rgw\_socket\_path if frontend is configured (#11160, Yehuda Sadeh)
- rgw: swift: Response header of GET request for container does not contain X-Container-Object-Count, X-Container-Bytes-Used and x-trans-id headers (#10666,

Dmytro Iurchenko

- rgw: swift: Response header of POST request for object does not contain content-length and x-trans-id headers (#10661, Radoslaw Zarzynski)
- rgw: swift: response for GET/HEAD on container does not contain the X-Timestamp header (#10938, Radoslaw Zarzynski)
- rgw: swift: response for PUT on /container does not contain the mandatory Content-Length header when FCGI is used (#11036, #10971, Radoslaw Zarzynski)
- rgw: swift: wrong handling of empty metadata on Swift container (#11088, Radoslaw Zarzynski)
- tests: TestFlatIndex.cc races with TestLFNIndex.cc (#11217, Xinze Chi)
- tests: ceph-helpers kill\_daemons fails when kill fails (#11398, Loic Dachary)

For more detailed information, see [the complete changelog](#).

## v0.94.1 Hammer

---

This bug fix release fixes a few critical issues with CRUSH. The most important addresses a bug in feature bit enforcement that may prevent pre-hammer clients from communicating with the cluster during an upgrade. This only manifests in some cases (for example, when the ‘rack’ type is in use in the CRUSH map, and possibly other cases), but for safety we strongly recommend that all users use 0.94.1 instead of 0.94 when upgrading.

There is also a fix in the new straw2 buckets when OSD weights are 0.

We recommend that all v0.94 users upgrade.

## Notable changes

---

- crush: fix divide-by-0 in straw2 (#11357 Sage Weil)
- crush: fix has\_v4\_buckets (#11364 Sage Weil)
- osd: fix negative degraded objects during backfilling (#7737 Guang Yang)

For more detailed information, see [the complete changelog](#).

# v0.94 Hammer

This major release is expected to form the basis of the next long-term stable series. It is intended to supersede v0.80.x Firefly.

Highlights since Giant include:

- *RADOS Performance*: a range of improvements have been made in the OSD and client-side librados code that improve the throughput on flash backends and improve parallelism and scaling on fast machines.
- *Simplified RGW deployment*: the ceph-deploy tool now has a new ‘ceph-deploy rgw create HOST’ command that quickly deploys a instance of the S3/Swift gateway using the embedded Civetweb server. This is vastly simpler than the previous Apache-based deployment. There are a few rough edges (e.g., around SSL support) but we encourage users to try [the new method](#).
- *RGW object versioning*: RGW now supports the S3 object versioning API, which preserves old version of objects instead of overwriting them.
- *RGW bucket sharding*: RGW can now shard the bucket index for large buckets across, improving performance for very large buckets.
- *RBD object maps*: RBD now has an object map function that tracks which parts of the image are allocating, improving performance for clones and for commands like export and delete.
- *RBD mandatory locking*: RBD has a new mandatory locking framework (still disabled by default) that adds additional safeguards to prevent multiple clients from using the same image at the same time.
- *RBD copy-on-read*: RBD now supports copy-on-read for image clones, improving performance for some workloads.
- *CephFS snapshot improvements*: Many many bugs have been fixed with CephFS snapshots. Although they are still disabled by default, stability has improved significantly.
- *CephFS Recovery tools*: We have built some journal recovery and diagnostic tools. Stability and performance of single-MDS systems is vastly improved in Giant, and more improvements have been made now in Hammer. Although we still recommend caution when storing important data in CephFS, we do encourage testing for non-critical workloads so that we can better gauge the feature, usability, performance, and stability gaps.
- *CRUSH improvements*: We have added a new straw2 bucket algorithm that reduces the amount of data migration required when changes are made to the cluster.

- *Shingled erasure codes (SHEC)*: The OSDs now have experimental support for shingled erasure codes, which allow a small amount of additional storage to be traded for improved recovery performance.
- *RADOS cache tiering*: A series of changes have been made in the cache tiering code that improve performance and reduce latency.
- *RDMA support*: There is now experimental support the RDMA via the Accelio (libxlio) library.
- *New administrator commands*: The ‘ceph osd df’ command shows pertinent details on OSD disk utilizations. The ‘ceph pg ls ...’ command makes it much simpler to query PG states while diagnosing cluster issues.

Other highlights since Firefly include:

- *CephFS*: we have fixed a raft of bugs in CephFS and built some basic journal recovery and diagnostic tools. Stability and performance of single-MDS systems is vastly improved in Giant. Although we do not yet recommend CephFS for production deployments, we do encourage testing for non-critical workloads so that we can better gauge the feature, usability, performance, and stability gaps.
- *Local Recovery Codes*: the OSDs now support an erasure-coding scheme that stores some additional data blocks to reduce the IO required to recover from single OSD failures.
- *Degraded vs misplaced*: the Ceph health reports from ‘ceph -s’ and related commands now make a distinction between data that is degraded (there are fewer than the desired number of copies) and data that is misplaced (stored in the wrong location in the cluster). The distinction is important because the latter does not compromise data safety.
- *Tiering improvements*: we have made several improvements to the cache tiering implementation that improve performance. Most notably, objects are not promoted into the cache tier by a single read; they must be found to be sufficiently hot before that happens.
- *Monitor performance*: the monitors now perform writes to the local data store asynchronously, improving overall responsiveness.
- *Recovery tools*: the ceph-objectstore-tool is greatly expanded to allow manipulation of an individual OSDs data store for debugging and repair purposes. This is most heavily used by our QA infrastructure to exercise recovery code.

I would like to take this opportunity to call out the amazing growth in contributors to Ceph beyond the core development team from Inktank. Hammer features major new features and improvements from Intel, Fujitsu, UnitedStack, Yahoo, UbuntuKylin, CohortFS, Mellanox, CERN, Deutsche Telekom, Mirantis, and SanDisk.

# Dedication

This release is dedicated in memoriam to Sandon Van Ness, aka Houkouonchi, who unexpectedly passed away a few weeks ago. Sandon was responsible for maintaining the large and complex Sepia lab that houses the Ceph project's build and test infrastructure. His efforts have made an important impact on our ability to reliably test Ceph with a relatively small group of people. He was a valued member of the team and we will miss him. H is also for Houkouonchi.

## Upgrading

- If your existing cluster is running a version older than v0.80.x Firefly, please first upgrade to the latest Firefly release before moving on to Giant. We have not tested upgrades directly from Emperor, Dumpling, or older releases.

We have tested:

- Firefly to Hammer
- Giant to Hammer
- Dumpling to Firefly to Hammer

- Please upgrade daemons in the following order:

- i. Monitors
- ii. OSDs
- iii. MDSS and/or radosgw

Note that the relative ordering of OSDs and monitors should not matter, but we primarily tested upgrading monitors first.

- The ceph-osd daemons will perform a disk-format upgrade improve the PG metadata layout and to repair a minor bug in the on-disk format. It may take a minute or two for this to complete, depending on how many objects are stored on the node; do not be alarmed if they do not marked "up" by the cluster immediately after starting.
- If upgrading from v0.93, set

```
osd enable degraded writes = false
```

on all osds prior to upgrading. The degraded writes feature has been reverted due to 11155.

- The LTTNG tracing in librbd and librados is disabled in the release packages until we find a way to avoid violating distro security policies when linking libust.

## Upgrading from v0.87.x Giant

- librbd and librados include ltng tracepoints on distros with libltng 2.4 or later (only Ubuntu Trusty for the ceph.com packages). When running a daemon that uses these libraries, i.e. an application that calls fork(2) or clone(2) without exec(3), you must set LD\_PRELOAD=libltng-ust-fork.so.0 to prevent a crash in the ltng atexit handler when the process exits. The only ceph tool that requires this is rbd-fuse.
- If rgw\_socket\_path is defined and rgw\_frontends defines a socket\_port and socket\_host, we now allow the rgw\_frontends settings to take precedence. This change should only affect users who have made non-standard changes to their radosgw configuration.
- If you are upgrading specifically from v0.92, you must stop all OSD daemons and flush their journals (`ceph-osd -i NNN --flush-journal`) before upgrading. There was a transaction encoding bug in v0.92 that broke compatibility. Upgrading from v0.93, v0.91, or anything earlier is safe.
- The experimental ‘keyvaluestore-dev’ OSD backend has been renamed ‘keyvaluestore’ (for simplicity) and marked as experimental. To enable this untested feature and acknowledge that you understand that it is untested and may destroy data, you need to add the following to your ceph.conf:

```
1. enable experimental unrecoverable data corrupting features = keyvaluestore
```

- The following librados C API function calls take a ‘flags’ argument whose value is now correctly interpreted:

```
rados_write_op_operate() rados_aio_write_op_operate() rados_read_op_operate() rados_aio_read_op_operate()
```

The flags were not correctly being translated from the librados constants to the internal values. Now they are. Any code that is passing flags to these methods should be audited to ensure that they are using the correct LIBRADOS\_OP\_FLAG\_\* constants.

- The ‘rados’ CLI ‘copy’ and ‘cppool’ commands now use the copy-from operation, which means the latest CLI cannot run these commands against pre-firefly OSDs.
- The librados watch/notify API now includes a watch\_flush() operation to flush the async queue of notify operations. This should be called by any watch/notify user prior to rados\_shutdown().

- The ‘category’ field for objects has been removed. This was originally added to track PG stat summations over different categories of objects for use by radosgw. It is no longer has any known users and is prone to abuse because it can lead to a pg\_stat\_t structure that is unbounded. The librados API calls that accept this field now ignore it, and the OSD no longer tracks the per-category summations.
- The output for ‘rados df’ has changed. The ‘category’ level has been eliminated, so there is now a single stat object per pool. The structure of the JSON output is different, and the plaintext output has one less column.
- The ‘rados create <objectname> [category]’ optional category argument is no longer supported or recognized.
- rados.py’s Rados class no longer has a `__del__` method; it was causing problems on interpreter shutdown and use of threads. If your code has Rados objects with limited lifetimes and you’re concerned about locked resources, call `Rados.shutdown()` explicitly.
- There is a new version of the librados watch/notify API with vastly improved semantics. Any applications using this interface are encouraged to migrate to the new API. The old API calls are marked as deprecated and will eventually be removed.
- The librados `rados_unwatch()` call used to be safe to call on an invalid handle. The new version has undefined behavior when passed a bogus value (for example, when `rados_watch()` returns an error and handle is not defined).
- The structure of the formatted ‘pg stat’ command is changed for the portion that counts states by name to avoid using the ‘+’ character (which appears in state names) as part of the XML token (it is not legal).
- Previously, the formatted output of ‘ceph pg stat -f ...’ was a full pg dump that included all metadata about all PGs in the system. It is now a concise summary of high-level PG stats, just like the unformatted ‘ceph pg stat’ command.
- All JSON dumps of floating point values were incorrectly surrounding the value with quotes. These quotes have been removed. Any consumer of structured JSON output that was consuming the floating point values was previously having to interpret the quoted string and will most likely need to be fixed to take the unquoted number.
- New ability to list all objects from all namespaces that can fail or return incomplete results when not all OSDs have been upgraded. Features `rados -all ls`, `rados cppool`, `rados export`, `rados cache-flush-evict-all` and `rados cache-try-flush-evict-all` can also fail or return incomplete results.
- Due to a change in the Linux kernel version 3.18 and the limits of the FUSE interface, `ceph-fuse` needs to be mounted as root on at least some systems. See issues #9997, #10277, and #10542 for details.

# Upgrading from v0.80x Firefly (additional notes)

- The client-side caching for librbd is now enabled by default (`rbd cache = true`). A safety option (`rbd cache writethrough until flush = true`) is also enabled so that writeback caching is not used until the library observes a ‘flush’ command, indicating that the librbd users is passing that operation through from the guest VM. This avoids potential data loss when used with older versions of qemu that do not support flush.

```
leveldb_write_buffer_size = 8*1024*1024 = 33554432 // 8MB leveldb_cache_size = 512*1024*1204 = 536870912 //
512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

- The ‘`rados getxattr ...`’ command used to add a gratuitous newline to the attr value; it now does not.
- The `*_kb perf` counters on the monitor have been removed. These are replaced with a new set of `*_bytes` counters (e.g., `cluster_osd_kb` is replaced by `cluster_osd_bytes`).
- The `rd_kb` and `wr_kb` fields in the JSON dumps for pool stats (accessed via the `ceph df detail -f json-pretty` and related commands) have been replaced with corresponding `*_bytes` fields. Similarly, the `total_space`, `total_used`, and `total_avail` fields are replaced with `total_bytes`, `total_used_bytes`, and `total_avail_bytes` fields.
- The `rados df --format=json` output `read_bytes` and `write_bytes` fields were incorrectly reporting ops; this is now fixed.
- The `rados df --format=json` output previously included `read_kb` and `write_kb` fields; these have been removed. Please use `read_bytes` and `write_bytes` instead (and divide by 1024 if appropriate).
- The experimental keyvaluestore-dev OSD backend had an on-disk format change that prevents existing OSD data from being upgraded. This affects developers and testers only.
- mon-specific and osd-specific leveldb options have been removed. From this point onward users should use the `leveldb_*` generic options and add the options in the appropriate sections of their configuration files. Monitors will still maintain the following monitor-specific defaults:

```
leveldb_write_buffer_size = 8*1024*1024 = 33554432 // 8MB leveldb_cache_size = 512*1024*1204 = 536870912 //
512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

- CephFS support for the legacy anchor table has finally been removed. Users with file systems created before firefly should ensure that inodes with multiple hard links are modified *prior* to the upgrade to ensure that the backtraces are written properly. For example:

```
1. sudo find /mnt/cephfs -type f -links +1 -exec touch \{\} \;
```

- We disallow nonsensical ‘tier cache-mode’ transitions. From this point onward, ‘writeback’ can only transition to ‘forward’ and ‘forward’ can transition to 1) ‘writeback’ if there are dirty objects, or 2) any if there are no dirty objects.

## Notable changes since v0.93

---

- build: a few cmake fixes (Matt Benjamin)
- build: fix build on RHEL/CentOS 5.9 (Rohan Mars)
- build: reorganize Makefile to allow modular builds (Boris Ranto)
- ceph-fuse: be more forgiving on remount (#10982 Greg Farnum)
- ceph: improve CLI parsing (#11093 David Zafman)
- common: fix cluster logging to default channel (#11177 Sage Weil)
- crush: fix parsing of straw2 buckets (#11015 Sage Weil)
- doc: update man pages (David Zafman)
- librados: fix leak in C\_TwoContexts (Xiong Yiliang)
- librados: fix leak in watch/notify path (Sage Weil)
- librbd: fix and improve AIO cache invalidation (#10958 Jason Dillaman)
- librbd: fix memory leak (Jason Dillaman)
- librbd: fix ordering/queueing of resize operations (Jason Dillaman)
- librbd: validate image is r/w on resize/flatten (Jason Dillaman)
- librbd: various internal locking fixes (Jason Dillaman)
- lttng: tracing is disabled until we streamline dependencies (Josh Durgin)
- mon: add bootstrap-rgw profile (Sage Weil)

- mon: do not pollute mon dir with CSV files from CRUSH check (Loic Dachary)
- mon: fix clock drift time check interval (#10546 Joao Eduardo Luis)
- mon: fix units in store stats (Joao Eduardo Luis)
- mon: improve error handling on erasure code profile set (#10488, #11144 Loic Dachary)
- mon: set {read,write}\_tier on 'osd tier add-cache ...' (Jianpeng Ma)
- ms: xio: fix misc bugs (Matt Benjamin, Vu Pham)
- osd: DBObjectMap: fix locking to prevent rare crash (#9891 Samuel Just)
- osd: fix and document last\_epoch\_started semantics (Samuel Just)
- osd: fix divergent entry handling on PG split (Samuel Just)
- osd: fix leak on shutdown (Kefu Chai)
- osd: fix recording of digest on scrub (Samuel Just)
- osd: fix whiteout handling (Sage Weil)
- rbd: allow v2 striping parameters for clones and imports (Jason Dillaman)
- rbd: fix formatted output of image features (Jason Dillaman)
- rbd: update eman page (Ilya Dryomov)
- rgw: don't overwrite bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: enable IPv6 for civetweb (#10965 Yehuda Sadeh)
- rgw: fix sysvinit script when rgw\_socket\_path is not defined (#11159 Yehuda Sadeh, Dan Mick)
- rgw: pass civetweb configurables through (#10907 Yehuda Sadeh)
- rgw: use new watch/notify API (Yehuda Sadeh, Sage Weil)
- osd: reverted degraded writes feature due to 11155

## Notable changes since v0.87.x Giant

---

- add experimental features option (Sage Weil)
- arch: fix NEON feaeture detection (#10185 Loic Dachary)
- asyncmsgr: misc fixes (Haomai Wang)
- buffer: add 'shareable' construct (Matt Benjamin)

- buffer: add list::get\_contiguous (Sage Weil)
- buffer: avoid rebuild if buffer already contiguous (Jianpeng Ma)
- build: CMake support (Ali Maredia, Casey Bodley, Adam Emerson, Marcus Watts, Matt Benjamin)
- build: a few cmake fixes (Matt Benjamin)
- build: aarch64 build fixes (Noah Watkins, Haomai Wang)
- build: adjust build deps for yasm, virtualenv (Jianpeng Ma)
- build: fix 'make check' races (#10384 Loic Dachary)
- build: fix build on RHEL/CentOS 5.9 (Rohan Mars)
- build: fix pkg names when libkeyutils is missing (Pankag Garg, Ken Dreyer)
- build: improve build dependency tooling (Loic Dachary)
- build: reorganize Makefile to allow modular builds (Boris Ranto)
- build: support for jemalloc (Shishir Gowda)
- ceph-disk: Scientific Linux support (Dan van der Ster)
- ceph-disk: allow journal partition re-use (#10146 Loic Dachary, Dav van der Ster)
- ceph-disk: call partx/partprobe consistency (#9721 Loic Dachary)
- ceph-disk: do not re-use partition if encryption is required (Loic Dachary)
- ceph-disk: fix dmcrypt key permissions (Loic Dachary)
- ceph-disk: fix umount race condition (#10096 Blaine Gardner)
- ceph-disk: improved systemd support (Owen Synge)
- ceph-disk: init=none option (Loic Dachary)
- ceph-disk: misc fixes (Christos Stavrakakis)
- ceph-disk: respect -statedir for keyring (Loic Dachary)
- ceph-disk: set guid if reusing journal partition (Dan van der Ster)
- ceph-disk: support LUKS for encrypted partitions (Andrew Bartlett, Loic Dachary)
- ceph-fuse, libcephfs: POSIX file lock support (Yan, Zheng)
- ceph-fuse, libcephfs: allow xattr caps in inject\_release\_failure (#9800 John Spray)

- ceph-fuse, libcephfs: fix I\_COMPLETE\_ORDERED checks (#9894 Yan, Zheng)
- ceph-fuse, libcephfs: fix cap flush overflow (Greg Farnum, Yan, Zheng)
- ceph-fuse, libcephfs: fix root inode xattrs (Yan, Zheng)
- ceph-fuse, libcephfs: preserve dir ordering (#9178 Yan, Zheng)
- ceph-fuse, libcephfs: trim inodes before reconnecting to MDS (Yan, Zheng)
- ceph-fuse, libcephfs: add support for O\_NOFOLLOW and O\_PATH (Greg Farnum)
- ceph-fuse, libcephfs: resend requests before completing cap reconnect (#10912 Yan, Zheng)
- ceph-fuse: be more forgiving on remount (#10982 Greg Farnum)
- ceph-fuse: fix dentry invalidation on 3.18+ kernels (#9997 Yan, Zheng)
- ceph-fuse: fix kernel cache trimming (#10277 Yan, Zheng)
- ceph-fuse: select kernel cache invalidation mechanism based on kernel version (Greg Farnum)
- ceph-monstore-tool: fix shutdown (#10093 Loic Dachary)
- ceph-monstore-tool: fix/improve CLI (Joao Eduardo Luis)
- ceph-objectstore-tool: fix import (#10090 David Zafman)
- ceph-objectstore-tool: improved import (David Zafman)
- ceph-objectstore-tool: many improvements and tests (David Zafman)
- ceph-objectstore-tool: many many improvements (David Zafman)
- ceph-objectstore-tool: misc improvements, fixes (#9870 #9871 David Zafman)
- ceph.spec: package rbd-replay-prep (Ken Dreyer)
- ceph: add 'ceph osd df [tree]' command (#10452 Mykola Golub)
- ceph: do not parse injectargs twice (Loic Dachary)
- ceph: fix 'ceph tell ...' command validation (#10439 Joao Eduardo Luis)
- ceph: improve 'ceph osd tree' output (Mykola Golub)
- ceph: improve CLI parsing (#11093 David Zafman)
- ceph: make 'ceph -s' output more readable (Sage Weil)
- ceph: make 'ceph -s' show PG state counts in sorted order (Sage Weil)

- ceph: make 'ceph tell mon.\* version' work (Mykola Golub)
- ceph: new 'ceph tell mds.\$name\_or\_rank\_or\_gid' (John Spray)
- ceph: show primary-affinity in 'ceph osd tree' (Mykola Golub)
- ceph: test robustness (Joao Eduardo Luis)
- ceph\_objectstore\_tool: behave with sharded flag (#9661 David Zafman)
- cephfs-journal-tool: add recover\_dentries function (#9883 John Spray)
- cephfs-journal-tool: fix journal import (#10025 John Spray)
- cephfs-journal-tool: skip up to expire\_pos (#9977 John Spray)
- cleanup rados.h definitions with macros (Ilya Dryomov)
- common: add 'perf reset ...' admin command (Jianpeng Ma)
- common: add TableFormatter (Andreas Peters)
- common: add newline to flushed json output (Sage Weil)
- common: check syncfs() return code (Jianpeng Ma)
- common: do not unlock rwlock on destruction (Federico Simoncelli)
- common: filtering for 'perf dump' (John Spray)
- common: fix Formatter factory breakage (#10547 Loic Dachary)
- common: fix block device discard check (#10296 Sage Weil)
- common: make json-pretty output prettier (Sage Weil)
- common: remove broken CEPH\_LOCKDEP optoin (Kefu Chai)
- common: shared\_cache unit tests (Cheng Cheng)
- common: support new gperftools header locations (Key Dreyer)
- config: add \$cctid meta variable (Adam Crume)
- crush: fix buffer overrun for poorly formed rules (#9492 Johnu George)
- crush: fix detach\_bucket (#10095 Sage Weil)
- crush: fix parsing of straw2 buckets (#11015 Sage Weil)
- crush: fix several bugs in adjust\_item\_weight (Rongze Zhu)
- crush: fix tree bucket behavior (Rongze Zhu)

- crush: improve constness (Loic Dachary)
- crush: new and improved straw2 bucket type (Sage Weil, Christina Anderson, Xiaoxi Chen)
- crush: straw bucket weight calculation fixes (#9998 Sage Weil)
- crush: update tries stats for indep rules (#10349 Loic Dachary)
- crush: use larger choose\_tries value for erasure code rulesets (#10353 Loic Dachary)
- crushtool: add -location <id> command (Sage Weil, Loic Dachary)
- debian,rpm: move RBD udev rules to ceph-common (#10864 Ken Dreyer)
- debian: split python-ceph into python-{rbd,rados,cephfs} (Boris Ranto)
- default to libnss instead of crypto++ (Federico Gimenez)
- doc: CephFS disaster recovery guidance (John Spray)
- doc: CephFS for early adopters (John Spray)
- doc: add build-doc guidlines for Fedora and CentOS/RHEL (Nilamdyuti Goswami)
- doc: add dumpling to firefly upgrade section (#7679 John Wilkins)
- doc: ceph osd reweight vs crush weight (Laurent Guerby)
- doc: do not suggest dangerous XFS nobarrier option (Dan van der Ster)
- doc: document erasure coded pool operations (#9970 Loic Dachary)
- doc: document the LRC per-layer plugin configuration (Yuan Zhou)
- doc: enable rbd cache on openstack deployments (Sebastien Han)
- doc: erasure code doc updates (Loic Dachary)
- doc: file system osd config settings (Kevin Dalley)
- doc: fix OpenStack Glance docs (#10478 Sebastien Han)
- doc: improved installation nots on CentOS/RHEL installs (John Wilkins)
- doc: key/value store config reference (John Wilkins)
- doc: misc cleanups (Adam Spiers, Sebastien Han, Nilamdyuti Goswami, Ken Dreyer, John Wilkins)
- doc: misc improvements (Nilamdyuti Goswami, John Wilkins, Chris Holcombe)
- doc: misc updates (#9793 #9922 #10204 #10203 Travis Rhoden, Hazem, Ayari, Florian

Coste, Andy Allan, Frank Yu, Baptiste Veuillez-Mainard, Yuan Zhou, Armando Segnini, Robert Jansen, Tyler Brekke, Viktor Suprun)

- doc: misc updates (Alfredo Deza, VRan Liu)
- doc: misc updates (Nilamdyuti Goswami, John Wilkins)
- doc: new man pages (Nilamdyuti Goswami)
- doc: preflight doc fixes (John Wilkins)
- doc: replace cloudfiles with swiftclient Python Swift example (Tim Freund)
- doc: update PG count guide (Gerben Meijer, Laurent Guerby, Loic Dachary)
- doc: update man pages (David Zafman)
- doc: update openstack docs for Juno (Sebastien Han)
- doc: update release descriptions (Ken Dreyer)
- doc: update sepia hardware inventory (Sandon Van Ness)
- erasure-code: add mSHEC erasure code support (Takeshi Miyamae)
- erasure-code: improved docs (#10340 Loic Dachary)
- erasure-code: set max\_size to 20 (#10363 Loic Dachary)
- fix cluster logging from non-mon daemons (Sage Weil)
- init-ceph: check for systemd-run before using it (Boris Ranto)
- install-deps.sh: do not require sudo when root (Loic Dachary)
- keyvaluestore: misc fixes (Haomai Wang)
- keyvaluestore: performance improvements (Haomai Wang)
- libcephfs,ceph-fuse: add 'status' asok (John Spray)
- libcephfs,ceph-fuse: fix getting zero-length xattr (#10552 Yan, Zheng)
- libcephfs: fix dirfrag trimming (#10387 Yan, Zheng)
- libcephfs: fix mount timeout (#10041 Yan, Zheng)
- libcephfs: fix test (#10415 Yan, Zheng)
- libcephfs: fix use-afer-free on umount (#10412 Yan, Zheng)
- libcephfs: include ceph and git version in client metadata (Sage Weil)
- librados, osd: new watch/notify implementation (Sage Weil)

- librados: add blacklist\_add convenience method (Jason Dillaman)
- librados: add rados\_pool\_get\_base\_tier() call (Adam Crume)
- librados: add watch\_flush() operation (Sage Weil, Haomai Wang)
- librados: avoid memcpy on getxattr, read (Jianpeng Ma)
- librados: cap buffer length (Loic Dachary)
- librados: create ioctx by pool id (Jason Dillaman)
- librados: do notify completion in fast-dispatch (Sage Weil)
- librados: drop 'category' feature (Sage Weil)
- librados: expose rados\_{read|write}\_op\_assert\_version in C API (Kim Vandry)
- librados: fix infinite loop with skipped map epochs (#9986 Ding Dinghua)
- librados: fix iterator operator= bugs (#10082 David Zafman, Yehuda Sadeh)
- librados: fix leak in C\_TwoContexts (Xiong Yiliang)
- librados: fix leak in watch/notify path (Sage Weil)
- librados: fix null deref when pool DNE (#9944 Sage Weil)
- librados: fix objecter races (#9617 Josh Durgin)
- librados: fix pool deletion handling (#10372 Sage Weil)
- librados: fix pool name caching (#10458 Radoslaw Zarzynski)
- librados: fix resource leak, misc bugs (#10425 Radoslaw Zarzynski)
- librados: fix some watch/notify locking (Jason Dillaman, Josh Durgin)
- librados: fix timer race from recent refactor (Sage Weil)
- librados: new fadvise API (Ma Jianpeng)
- librados: only export public API symbols (Jason Dillaman)
- librados: remove shadowed variable (Kefu Chain)
- librados: translate op flags from C APIs (Matthew Richards)
- libradosstriper: fix remove() (Dongmao Zhang)
- libradosstriper: fix shutdown hang (Dongmao Zhang)
- libradosstriper: fix stat strtoll (Dongmao Zhang)

- libradosstriper: fix trunc method (#10129 Sebastien Ponce)
- libradosstriper: fix write\_full when ENOENT (#10758 Sebastien Ponce)
- libradosstriper: misc fixes (Sebastien Ponce)
- librbd: CRC protection for RBD image map (Jason Dillaman)
- librbd: add missing python docstrings (Jason Dillaman)
- librbd: add per-image object map for improved performance (Jason Dillaman)
- librbd: add readahead (Adam Crume)
- librbd: add support for an “object map” indicating which objects exist (Jason Dillaman)
- librbd: adjust internal locking (Josh Durgin, Jason Dillaman)
- librbd: better handling of watch errors (Jason Dillaman)
- librbd: complete pending ops before closing image (#10299 Josh Durgin)
- librbd: coordinate maint operations through lock owner (Jason Dillaman)
- librbd: copy-on-read (Min Chen, Li Wang, Yunchuan Wen, Cheng Cheng, Jason Dillaman)
- librbd: differentiate between R/O vs R/W features (Jason Dillaman)
- librbd: don't close a closed parent in failure path (#10030 Jason Dillaman)
- librbd: enforce write ordering with a snapshot (Jason Dillaman)
- librbd: exclusive image locking (Jason Dillaman)
- librbd: fadvise API (Ma Jianpeng)
- librbd: fadvise-style hints; add misc hints for certain operations (Jianpeng Ma)
- librbd: fix and improve AIO cache invalidation (#10958 Jason Dillaman)
- librbd: fix cache tiers in list\_children and snap\_unprotect (Adam Crume)
- librbd: fix coverity false-positives (Jason Dillaman)
- librbd: fix diff test (#10002 Josh Durgin)
- librbd: fix list\_children from invalid pool ioctxs (#10123 Jason Dillaman)
- librbd: fix locking for readahead (#10045 Jason Dillaman)
- librbd: fix memory leak (Jason Dillaman)

- librbd: fix ordering/queueing of resize operations (Jason Dillaman)
- librbd: fix performance regression in ObjectCacher (#9513 Adam Crume)
- librbd: fix snap create races (Jason Dillaman)
- librbd: fix write vs import race (#10590 Jason Dillaman)
- librbd: flush AIO operations asynchronously (#10714 Jason Dillaman)
- librbd: gracefully handle deleted/renamed pools (#10270 Jason Dillaman)
- librbd: lttng tracepoints (Adam Crume)
- librbd: make async versions of long-running maint operations (Jason Dillaman)
- librbd: misc fixes (Xinxin Shu, Jason Dillaman)
- librbd: mock tests (Jason Dillaman)
- librbd: only export public API symbols (Jason Dillaman)
- librbd: optionally blacklist clients before breaking locks (#10761 Jason Dillaman)
- librbd: prevent copyup during shrink (Jason Dillaman)
- librbd: refactor unit tests to use fixtures (Jason Dillaman)
- librbd: validate image is r/w on resize/flatten (Jason Dillaman)
- librbd: various internal locking fixes (Jason Dillaman)
- many coverity fixes (Danny Al-Gaaf)
- many many coverity cleanups (Danny Al-Gaaf)
- mds: 'flush journal' admin command (John Spray)
- mds: ENOSPC and OSDMap epoch barriers (#7317 John Spray)
- mds: a whole bunch of initial scrub infrastructure (Greg Farnum)
- mds: add cephfs-table-tool (John Spray)
- mds: asok command for fetching subtree map (John Spray)
- mds: avoid sending traceless replies in most cases (Yan, Zheng)
- mds: constify MDSCacheObjects (John Spray)
- mds: dirfrag buf fix (Yan, Zheng)
- mds: disallow most commands on inactive MDS's (Greg Farnum)

- mds: drop dentries, leases on deleted directories (#10164 Yan, Zheng)
- mds: export dir asok command (John Spray)
- mds: fix MDLog IO callback deadlock (John Spray)
- mds: fix compat\_version for MClientSession (#9945 John Spray)
- mds: fix deadlock during journal probe vs purge (#10229 Yan, Zheng)
- mds: fix race trimming log segments (Yan, Zheng)
- mds: fix reply snapbl (Yan, Zheng)
- mds: fix sessionmap lifecycle bugs (Yan, Zheng)
- mds: fix stray/purge perfcounters (#10388 John Spray)
- mds: handle heartbeat\_reset during shutdown (#10382 John Spray)
- mds: handle zero-size xattr (#10335 Yan, Zheng)
- mds: initialize root inode xattr version (Yan, Zheng)
- mds: introduce auth caps (John Spray)
- mds: many many snapshot-related fixes (Yan, Zheng)
- mds: misc bugs (Greg Farnum, John Spray, Yan, Zheng, Henry Change)
- mds: refactor, improve Session storage (John Spray)
- mds: store backtrace for stray dir (Yan, Zheng)
- mds: subtree quota support (Yunchuan Wen)
- mds: verify backtrace when fetching dirfrag (#9557 Yan, Zheng)
- memstore: free space tracking (John Spray)
- misc cleanup (Danny Al-Gaaf, David Anderson)
- misc coverity fixes (Danny Al-Gaaf)
- misc coverity fixes (Danny Al-Gaaf)
- misc: various valgrind fixes and cleanups (Danny Al-Gaaf)
- mon: ‘osd crush reweight-all’ command (Sage Weil)
- mon: add ‘ceph osd rename-bucket ...’ command (Loic Dachary)
- mon: add bootstrap-rgw profile (Sage Weil)

- mon: add max pgs per osd warning (Sage Weil)
- mon: add noforward flag for some mon commands (Mykola Golub)
- mon: allow adding tiers to fs pools (#10135 John Spray)
- mon: allow full flag to be manually cleared (#9323 Sage Weil)
- mon: clean up auth list output (Loic Dachary)
- mon: delay failure injection (Joao Eduardo Luis)
- mon: disallow empty pool names (#10555 Wido den Hollander)
- mon: do not deactivate last mds (#10862 John Spray)
- mon: do not pollute mon dir with CSV files from CRUSH check (Loic Dachary)
- mon: drop old ceph\_mon\_store\_converter (Sage Weil)
- mon: fix ‘ceph pg dump\_stuck degraded’ (Xinxin Shu)
- mon: fix ‘mds fail’ for standby MDSs (John Spray)
- mon: fix ‘osd crush link’ id resolution (John Spray)
- mon: fix ‘profile osd’ use of config-key function on mon (#10844 Joao Eduardo Luis)
- mon: fix \_ratio units and types (Sage Weil)
- mon: fix JSON dumps to dump floats as floats and not strings (Sage Weil)
- mon: fix MDS health status from peons (#10151 John Spray)
- mon: fix caching for min\_last\_epoch\_clean (#9987 Sage Weil)
- mon: fix clock drift time check interval (#10546 Joao Eduardo Luis)
- mon: fix compatset initialization during mkfs (Joao Eduardo Luis)
- mon: fix error output for add\_data\_pool (#9852 Joao Eduardo Luis)
- mon: fix feature tracking during elections (Joao Eduardo Luis)
- mon: fix formatter ‘pg stat’ command output (Sage Weil)
- mon: fix mds gid/rank/state parsing (John Spray)
- mon: fix misc error paths (Joao Eduardo Luis)
- mon: fix paxos off-by-one corner case (#9301 Sage Weil)
- mon: fix paxos timeouts (#10220 Joao Eduardo Luis)

- mon: fix stashed monmap encoding (#5203 Xie Rui)
- mon: fix units in store stats (Joao Eduardo Luis)
- mon: get canonical OSDMap from leader (#10422 Sage Weil)
- mon: ignore failure reports from before up\_from (#10762 Dan van der Ster, Sage Weil)
- mon: implement 'fs reset' command (John Spray)
- mon: improve error handling on erasure code profile set (#10488, #11144 Loic Dachary)
- mon: improved corrupt CRUSH map detection (Joao Eduardo Luis)
- mon: include entity name in audit log for forwarded requests (#9913 Joao Eduardo Luis)
- mon: include pg\_temp count in osdmap summary (Sage Weil)
- mon: log health summary to cluster log (#9440 Joao Eduardo Luis)
- mon: make 'mds fail' idempotent (John Spray)
- mon: make pg dump {sum,pgs,pgs\_brief} work for format=plain (#5963 #6759 Mykola Golub)
- mon: new 'ceph pool ls [detail]' command (Sage Weil)
- mon: new pool safety flags nodelete, nopgchange, nosizechange (#9792 Mykola Golub)
- mon: new, friendly 'ceph pg ls ...' command (Xinxin Shu)
- mon: paxos: allow reads while proposing (#9321 #9322 Joao Eduardo Luis)
- mon: prevent MDS transition from STOPPING (#10791 Greg Farnum)
- mon: propose all pending work in one transaction (Sage Weil)
- mon: remove pg\_temps for nonexistent pools (Joao Eduardo Luis)
- mon: require mon\_allow\_pool\_delete option to remove pools (Sage Weil)
- mon: respect down flag when promoting standbys (John Spray)
- mon: set globalid prealloc to larger value (Sage Weil)
- mon: set {read,write}\_tier on 'osd tier add-cache ...' (Jianpeng Ma)
- mon: skip zeroed osd stats in get\_rule\_avail (#10257 Joao Eduardo Luis)

- mon: validate min\_size range (Jianpeng Ma)
- mon: wait for writeable before cross-proposing (#9794 Joao Eduardo Luis)
- mount.ceph: fix suprious error message (#10351 Yan, Zheng)
- ms: xio: fix misc bugs (Matt Benjamin, Vu Pham)
- msgr: async: bind threads to CPU cores, improved poll (Haomai Wang)
- msgr: async: many fixes, unit tests (Haomai Wang)
- msgr: async: several fixes (Haomai Wang)
- msgr: asyncmessenger: add kqueue support (#9926 Haomai Wang)
- msgr: avoid useless new/delete (Haomai Wang)
- msgr: fix RESETSESSION bug (#10080 Greg Farnum)
- msgr: fix crc configuration (Mykola Golub)
- msgr: fix delay injection bug (#9910 Sage Weil, Greg Farnum)
- msgr: misc unit tests (Haomai Wang)
- msgr: new AsymcMessenger alternative implementation (Haomai Wang)
- msgr: prefetch data when doing recv (Yehuda Sadeh)
- msgr: simple: fix rare deadlock (Greg Farnum)
- msgr: simple: retry binding to port on failure (#10029 Wido den Hollander)
- msgr: xio: XioMessenger RDMA support (Casey Bodley, Vu Pham, Matt Benjamin)
- objectstore: deprecate collection attrs (Sage Weil)
- osd, librados: fadvise-style librados hints (Jianpeng Ma)
- osd, librados: fix xattr\_cmp\_u64 (Dongmao Zhang)
- osd, librados: revamp PG listing API to handle namespaces (#9031 #9262 #9438 David Zafman)
- osd, mds: 'ops' as shorthand for 'dump\_ops\_in\_flight' on asok (Sage Weil)
- osd, mon: add checksums to all OSDMaps (Sage Weil)
- osd, mon: send intitial pg create time from mon to osd (#9887 David Zafman)
- osd,mon: add 'norebalance' flag (Kefu Chai)
- osd,mon: specify OSD features explicitly in MOSDBoot (#10911 Sage Weil)

- osd: DBObjectMap: fix locking to prevent rare crash (#9891 Samuel Just)
- osd: EIO on whole-object reads when checksum is wrong (Sage Weil)
- osd: add erasure code corpus (Loic Dachary)
- osd: add fadvise flags to ObjectStore API (Jianpeng Ma)
- osd: add get\_latest\_osdmap asok command (#9483 #9484 Mykola Golub)
- osd: add misc tests (Loic Dachary, Danny Al-Gaaf)
- osd: add option to prioritize heartbeat network traffic (Jian Wen)
- osd: add support for the SHEC erasure-code algorithm (Takeshi Miyamae, Loic Dachary)
- osd: allow deletion of objects with watcher (#2339 Sage Weil)
- osd: allow recovery while below min\_size (Samuel Just)
- osd: allow recovery with fewer than min\_size OSDs (Samuel Just)
- osd: allow sparse read for Push/Pull (Haomai Wang)
- osd: allow whiteout deletion in cache pool (Sage Weil)
- osd: allow writes to degraded objects (Samuel Just)
- osd: allow writes to degraded objects (Samuel Just)
- osd: avoid publishing unchanged PG stats (Sage Weil)
- osd: batch pg log trim (Xinze Chi)
- osd: cache pool: ignore min flush age when cache is full (Xinze Chi)
- osd: cache recent ObjectContexts (Dong Yuan)
- osd: cache reverse\_nibbles hash value (Dong Yuan)
- osd: clean up internal ObjectStore interface (Sage Weil)
- osd: cleanup boost optionals (William Kennington)
- osd: clear cache on interval change (Samuel Just)
- osd: do no proxy reads unless target OSDs are new (#10788 Sage Weil)
- osd: do not abort deep scrub on missing hinfo (#10018 Loic Dachary)
- osd: do not update digest on inconsistent object (#10524 Samuel Just)
- osd: don't record digests for snapdirs (#10536 Samuel Just)

- osd: drop upgrade support for pre-dumpling (Sage Weil)
- osd: enable and use posix\_fadvise (Sage Weil)
- osd: erasure coding: allow bench.sh to test ISA backend (Yuan Zhou)
- osd: erasure-code: encoding regression tests, corpus (#9420 Loic Dachary)
- osd: erasure-code: enforce chunk size alignment (#10211 Loic Dachary)
- osd: erasure-code: jerasure support for NEON (Loic Dachary)
- osd: erasure-code: relax cauchy w restrictions (#10325 David Zhang, Loic Dachary)
- osd: erasure-code: update gf-complete to latest upstream (Loic Dachary)
- osd: expose non-journal backends via ceph-osd CLI (Hoamai Wang)
- osd: filejournal: don't cache journal when not using direct IO (Jianpeng Ma)
- osd: fix JSON output for stray OSDs (Loic Dachary)
- osd: fix OSDCap parser on old (el6) boost::spirit (#10757 Kefu Chai)
- osd: fix OSDCap parsing on el6 (#10757 Kefu Chai)
- osd: fix ObjectStore::Transaction encoding version (#10734 Samuel Just)
- osd: fix WBTHrottle perf counters (Haomai Wang)
- osd: fix and document last\_epoch\_started semantics (Samuel Just)
- osd: fix auth object selection during repair (#10524 Samuel Just)
- osd: fix backfill bug (#10150 Samuel Just)
- osd: fix bug in pending digest updates (#10840 Samuel Just)
- osd: fix cancel\_proxy\_read\_ops (Sage Weil)
- osd: fix cleanup of interrupted pg deletion (#10617 Sage Weil)
- osd: fix divergent entry handling on PG split (Samuel Just)
- osd: fix ghobject\_t formatted output to include shard (#10063 Loic Dachary)
- osd: fix ioprio option (Mykola Golub)
- osd: fix ioprio options (Loic Dachary)
- osd: fix journal shutdown race (Sage Weil)
- osd: fix journal wrapping bug (#10883 David Zafman)

- osd: fix leak in SnapTrimWQ (#10421 Kefu Chai)
- osd: fix leak on shutdown (Kefu Chai)
- osd: fix memstore free space calculation (Xiaoxi Chen)
- osd: fix mixed-version peering issues (Samuel Just)
- osd: fix object age eviction (Zhiqiang Wang)
- osd: fix object atime calculation (Xinze Chi)
- osd: fix object digest update bug (#10840 Samuel Just)
- osd: fix occasional peering stalls (#10431 Sage Weil)
- osd: fix ordering issue with new transaction encoding (#10534 Dong Yuan)
- osd: fix osd peer check on scrub messages (#9555 Sage Weil)
- osd: fix past\_interval display bug (#9752 Loic Dachary)
- osd: fix past\_interval generation (#10427 #10430 David Zafman)
- osd: fix pgls filter ops (#9439 David Zafman)
- osd: fix recording of digest on scrub (Samuel Just)
- osd: fix scrub delay bug (#10693 Samuel Just)
- osd: fix scrub vs try-flush bug (#8011 Samuel Just)
- osd: fix short read handling on push (#8121 David Zafman)
- osd: fix stderr with -f or -d (Dan Mick)
- osd: fix transaction accounting (Jianpeng Ma)
- osd: fix watch reconnect race (#10441 Sage Weil)
- osd: fix watch timeout cache state update (#10784 David Zafman)
- osd: fix whiteout handling (Sage Weil)
- osd: flush snapshots from cache tier immediately (Sage Weil)
- osd: force promotion of watch/notify ops (Zhiqiang Wang)
- osd: handle no-op write with snapshot (#10262 Sage Weil)
- osd: improve idempotency detection across cache promotion/demotion (#8935 Sage Weil, Samuel Just)
- osd: include activating peers in blocked\_by (#10477 Sage Weil)

- osd: jerasure and gf-complete updates from upstream (#10216 Loic Dachary)
- osd: journal: check fsync/fdatasync result (Jianpeng Ma)
- osd: journal: fix alignment checks, avoid useless memmove (Jianpeng Ma)
- osd: journal: fix hang on shutdown (#10474 David Zafman)
- osd: journal: fix header.committed\_up\_to (Xinze Chi)
- osd: journal: fix journal zeroing when direct IO is enabled (Xie Rui)
- osd: journal: initialize throttle (Ning Yao)
- osd: journal: misc bug fixes (#6003 David Zafman, Samuel Just)
- osd: journal: update committed\_thru after replay (#6756 Samuel Just)
- osd: keyvaluestore: cleanup dead code (Ning Yao)
- osd: keyvaluestore: fix getattr semantics (Haomai Wang)
- osd: keyvaluestore: fix key ordering (#10119 Haomai Wang)
- osd: keyvaluestore\_dev: optimization (Chendi Xue)
- osd: limit in-flight read requests (Jason Dillaman)
- osd: log when scrub or repair starts (Loic Dachary)
- osd: make misdirected op checks robust for EC pools (#9835 Sage Weil)
- osd: memstore: fix size limit (Xiaoxi Chen)
- osd: misc FIEMAP fixes (Ma Jianpeng)
- osd: misc cleanup (Xinze Chi, Yongyue Sun)
- osd: misc optimizations (Xinxin Shu, Zhiqiang Wang, Xinze Chi)
- osd: misc scrub fixes (#10017 Loic Dachary)
- osd: new ‘activating’ state between peering and active (Sage Weil)
- osd: new optimized encoding for ObjectStore::Transaction (Dong Yuan)
- osd: optimize Finisher (Xinze Chi)
- osd: optimize WBThrottle map with unordered\_map (Ning Yao)
- osd: optimize filter\_snapc (Ning Yao)
- osd: preserve reqids for idempotency checks for promote/demote (Sage Weil, Zhiqiang Wang, Samuel Just)

- osd: proxy read support (Zhiqiang Wang)
- osd: proxy reads during cache promote (Zhiqiang Wang)
- osd: remove dead locking code (Xinxin Shu)
- osd: remove legacy classic scrub code (Sage Weil)
- osd: remove unused fields in MOSDSubOp (Xiaoxi Chen)
- osd: removed some dead code (Xinze Chi)
- osd: replace MOSDSubOp messages with simpler, optimized MOSDRepOp (Xiaoxi Chen)
- osd: restrict scrub to certain times of day (Xinze Chi)
- osd: rocksdb: fix shutdown (Hoamai Wang)
- osd: store PG metadata in per-collection objects for better concurrency (Sage Weil)
- osd: store whole-object checksums on scrub, write\_full (Sage Weil)
- osd: support for discard for journal trim (Jianpeng Ma)
- osd: use FIEMAP\_FLAGS\_SYNC instead of fsync (Jianpeng Ma)
- osd: verify kernel is new enough before using XFS extsize ioctl, enable by default (#9956 Sage Weil)
- pybind: fix memory leak in librados bindings (Billy Olsen)
- pyrados: add object lock support (#6114 Mehdi Abaakouk)
- pyrados: fix misnamed wait\_\* routings (#10104 Dan Mick)
- pyrados: misc cleanups (Kefu Chai)
- qa: add large auth ticket tests (Ilya Dryomov)
- qa: fix mds tests (#10539 John Spray)
- qa: fix osd create dup tests (#10083 Loic Dachary)
- qa: ignore duplicates in rados ls (Josh Durgin)
- qa: improve hadoop tests (Noah Watkins)
- qa: many 'make check' improvements (Loic Dachary)
- qa: misc tests (Loic Dachary, Yan, Zheng)
- qa: parallelize make check (Loic Dachary)

- qa: reorg fs quota tests (Greg Farnum)
- qa: tolerate nearly-full disk for make check (Loic Dachary)
- rados: fix put of /dev/null (Loic Dachary)
- rados: fix usage (Jianpeng Ma)
- rados: parse command-line arguments more strictly (#8983 Adam Crume)
- rados: use copy-from operation for copy, cppool (Sage Weil)
- radosgw-admin: add replicalog update command (Yehuda Sadeh)
- rbd-fuse: clean up on shutdown (Josh Durgin)
- rbd-fuse: fix memory leak (Adam Crume)
- rbd-replay-many (Adam Crume)
- rbd-replay: --anonymize flag to rbd-replay-prep (Adam Crume)
- rbd: add 'merge-diff' function (MingXin Liu, Yunchuan Wen, Li Wang)
- rbd: allow v2 striping parameters for clones and imports (Jason Dillaman)
- rbd: fix 'rbd diff' for non-existent objects (Adam Crume)
- rbd: fix buffer handling on image import (#10590 Jason Dillaman)
- rbd: fix error when striping with format 1 (Sebastien Han)
- rbd: fix export for image sizes over 2GB (Vicente Cheng)
- rbd: fix formatted output of image features (Jason Dillaman)
- rbd: leave exclusive lockin goff by default (Jason Dillaman)
- rbd: update eman page (Ilya Dryomov)
- rbd: update init-rbdmap to fix dup mount point (Karel Striegel)
- rbd: use IO hints for import, export, and bench operations (#10462 Jason Dillaman)
- rbd: use rolling average for rbd bench-write throughput (Jason Dillaman)
- rbd\_recover\_tool: RBD image recovery tool (Min Chen)
- rgw: S3-style object versioning support (Yehuda Sadeh)
- rgw: add location header when object is in another region (VRan Liu)
- rgw: change multipart upload id magic (#10271 Yehuda Sadeh)

- rgw: check keystone auth for S3 POST requests (#10062 Abhishek Lekshmanan)
- rgw: check timestamp on s3 keystone auth (#10062 Abhishek Lekshmanan)
- rgw: conditional PUT on ETag (#8562 Ray Lv)
- rgw: create subuser if needed when creating user (#10103 Yehuda Sadeh)
- rgw: decode http query params correction (#10271 Yehuda Sadeh)
- rgw: don't overwrite bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: enable IPv6 for civetweb (#10965 Yehuda Sadeh)
- rgw: extend replica log API (purge-all) (Yehuda Sadeh)
- rgw: fail S3 POST if keystone not configured (#10688 Valery Tschopp, Yehuda Sadeh)
- rgw: fix If-Modified-Since (VRan Liu)
- rgw: fix XML header on get ACL request (#10106 Yehuda Sadeh)
- rgw: fix bucket removal with data purge (Yehuda Sadeh)
- rgw: fix content length check (#10701 Axel Dunkel, Yehuda Sadeh)
- rgw: fix content-length update (#9576 Yehuda Sadeh)
- rgw: fix disabling of max\_size quota (#9907 Dong Lei)
- rgw: fix error codes (#10334 #10329 Yehuda Sadeh)
- rgw: fix incorrect len when len is 0 (#9877 Yehuda Sadeh)
- rgw: fix object copy content type (#9478 Yehuda Sadeh)
- rgw: fix partial GET in swift (#10553 Yehuda Sadeh)
- rgw: fix replica log indexing (#8251 Yehuda Sadeh)
- rgw: fix shutdown (#10472 Yehuda Sadeh)
- rgw: fix swift metadata header name (Dmytro Iurchenko)
- rgw: fix sysvinit script when rgw\_socket\_path is not defined (#11159 Yehuda Sadeh, Dan Mick)
- rgw: fix user stags in get-user-info API (#9359 Ray Lv)
- rgw: include XML ns on get ACL request (#10106 Yehuda Sadeh)
- rgw: index swift keys appropriately (#10471 Yehuda Sadeh)

- rgw: make sysvinit script set ulimit -n properly (Sage Weil)
- rgw: misc fixes (#10307 Yehuda Sadeh)
- rgw: only track cleanup for objects we write (#10311 Yehuda Sadeh)
- rgw: pass civetweb configurables through (#10907 Yehuda Sadeh)
- rgw: prevent illegal bucket policy that doesn't match placement rule (Yehuda Sadeh)
- rgw: remove multipart entries from bucket index on abort (#10719 Yehuda Sadeh)
- rgw: remove swift user manifest (DLO) hash calculation (#9973 Yehuda Sadeh)
- rgw: respond with 204 to POST on containers (#10667 Yuan Zhou)
- rgw: return timestamp on GET/HEAD (#8911 Yehuda Sadeh)
- rgw: reuse fcgx connection struct (#10194 Yehuda Sadeh)
- rgw: run radosgw as apache with systemd (#10125 Loic Dachary)
- rgw: send explicit HTTP status string (Yehuda Sadeh)
- rgw: set ETag on object copy (#9479 Yehuda Sadeh)
- rgw: set length for keystone token validation request (#7796 Yehuda Sadeh, Mark Kirkwood)
- rgw: support X-Storage-Policy header for Swift storage policy compat (Yehuda Sadeh)
- rgw: support multiple host names (#7467 Yehuda Sadeh)
- rgw: swift: dump container's custom metadata (#10665 Ahmad Faheem, Dmytro Iurchenko)
- rgw: swift: support Accept header for response format (#10746 Dmytro Iurchenko)
- rgw: swift: support for X-Remove-Container-Meta-{key} (#10475 Dmytro Iurchenko)
- rgw: tweak error codes (#10329 #10334 Yehuda Sadeh)
- rgw: update bucket index on attr changes, for multi-site sync (#5595 Yehuda Sadeh)
- rgw: use rn for http headers (#9254 Yehuda Sadeh)
- rgw: use gc for multipart abort (#10445 Aaron Bassett, Yehuda Sadeh)
- rgw: use new watch/notify API (Yehuda Sadeh, Sage Weil)

- rpm: misc fixes (Key Dreyer)
- rpm: move rgw logrotate to radosgw subpackage (Ken Dreyer)
- systemd: better systemd unit files (Owen Synge)
- sysvinit: fix race in ‘stop’ (#10389 Loic Dachary)
- test: fix bufferlist tests (Jianpeng Ma)
- tests: ability to run unit tests under docker (Loic Dachary)
- tests: centos-6 dockerfile (#10755 Loic Dachary)
- tests: improve docker-based tests (Loic Dachary)
- tests: unit tests for shared\_cache (Dong Yuan)
- udev: fix rules for CentOS7/RHEL7 (Loic Dachary)
- use clock\_gettime instead of gettimeofday (Jianpeng Ma)
- vstart.sh: set up environment for s3-tests (Luis Pabon)
- vstart.sh: work with cmake (Yehuda Sadeh)

## v0.93

---

This is the first release candidate for Hammer, and includes all of the features that will be present in the final release. We welcome and encourage any and all testing in non-production clusters to identify any problems with functionality, stability, or performance before the final Hammer release.

We suggest some caution in one area: librbd. There is a lot of new functionality around object maps and locking that is disabled by default but may still affect stability for existing images. We are continuing to shake out those bugs so that the final Hammer release (probably v0.94) will be rock solid.

Major features since Giant include:

- cephfs: journal scavenger repair tool (John Spray)
- crush: new and improved straw2 bucket type (Sage Weil, Christina Anderson, Xiaoxi Chen)
- doc: improved guidance for CephFS early adopters (John Spray)
- librbd: add per-image object map for improved performance (Jason Dillaman)
- librbd: copy-on-read (Min Chen, Li Wang, Yunchuan Wen, Cheng Cheng)

- librados: fadvise-style IO hints (Jianpeng Ma)
- mds: many many snapshot-related fixes (Yan, Zheng)
- mon: new ‘ceph osd df’ command (Mykola Golub)
- mon: new ‘ceph pg ls ...’ command (Xinxin Shu)
- osd: improved performance for high-performance backends
- osd: improved recovery behavior (Samuel Just)
- osd: improved cache tier behavior with reads (Zhiqiang Wang)
- rgw: S3-compatible bucket versioning support (Yehuda Sadeh)
- rgw: large bucket index sharding (Guang Yang, Yehuda Sadeh)
- RDMA “xio” messenger support (Matt Benjamin, Vu Pham)

## Upgrading

---

- If you are upgrading from v0.92, you must stop all OSD daemons and flush their journals ( `ceph-osd -i NNN --flush-journal` ) before upgrading. There was a transaction encoding bug in v0.92 that broke compatibility. Upgrading from v0.91 or anything earlier is safe.
- No special restrictions when upgrading from firefly or giant.

## Notable Changes

---

- build: CMake support (Ali Maredia, Casey Bodley, Adam Emerson, Marcus Watts, Matt Benjamin)
- ceph-disk: do not re-use partition if encryption is required (Loic Dachary)
- ceph-disk: support LUKS for encrypted partitions (Andrew Bartlett, Loic Dachary)
- ceph-fuse,libcephfs: add support for O\_NOFOLLOW and O\_PATH (Greg Farnum)
- ceph-fuse,libcephfs: resend requests before completing cap reconnect (#10912 Yan, Zheng)
- ceph-fuse: select kernel cache invalidation mechanism based on kernel version (Greg Farnum)
- ceph-objectstore-tool: improved import (David Zafman)
- ceph-objectstore-tool: misc improvements, fixes (#9870 #9871 David Zafman)

- ceph: add 'ceph osd df [tree]' command (#10452 Mykola Golub)
- ceph: fix 'ceph tell ...' command validation (#10439 Joao Eduardo Luis)
- ceph: improve 'ceph osd tree' output (Mykola Golub)
- cephfs-journal-tool: add recover\_dentries function (#9883 John Spray)
- common: add newline to flushed json output (Sage Weil)
- common: filtering for 'perf dump' (John Spray)
- common: fix Formatter factory breakage (#10547 Loic Dachary)
- common: make json-pretty output prettier (Sage Weil)
- crush: new and improved straw2 bucket type (Sage Weil, Christina Anderson, Xiaoxi Chen)
- crush: update tries stats for indep rules (#10349 Loic Dachary)
- crush: use larger choose\_tries value for erasure code rulesets (#10353 Loic Dachary)
- debian,rpm: move RBD udev rules to ceph-common (#10864 Ken Dreyer)
- debian: split python-ceph into python-{rbd,rados,cephfs} (Boris Ranto)
- doc: CephFS disaster recovery guidance (John Spray)
- doc: CephFS for early adopters (John Spray)
- doc: fix OpenStack Glance docs (#10478 Sebastien Han)
- doc: misc updates (#9793 #9922 #10204 #10203 Travis Rhoden, Hazem, Ayari, Florian Coste, Andy Allan, Frank Yu, Baptiste Veuillez-Mainard, Yuan Zhou, Armando Segnini, Robert Jansen, Tyler Brekke, Viktor Suprun)
- doc: replace cloudfiles with swiftclient Python Swift example (Tim Freund)
- erasure-code: add mSHEC erasure code support (Takeshi Miyamae)
- erasure-code: improved docs (#10340 Loic Dachary)
- erasure-code: set max\_size to 20 (#10363 Loic Dachary)
- libcephfs,ceph-fuse: fix getting zero-length xattr (#10552 Yan, Zheng)
- librados: add blacklist\_add convenience method (Jason Dillaman)
- librados: expose rados\_{read|write}\_op\_assert\_version in C API (Kim Vandry)
- librados: fix pool name caching (#10458 Radoslaw Zarzynski)

- librados: fix resource leak, misc bugs (#10425 Radoslaw Zarzynski)
- librados: fix some watch/notify locking (Jason Dillaman, Josh Durgin)
- libradosstriper: fix write\_full when ENOENT (#10758 Sebastien Ponce)
- librbd: CRC protection for RBD image map (Jason Dillaman)
- librbd: add per-image object map for improved performance (Jason Dillaman)
- librbd: add support for an “object map” indicating which objects exist (Jason Dillaman)
- librbd: adjust internal locking (Josh Durgin, Jason Dillaman)
- librbd: better handling of watch errors (Jason Dillaman)
- librbd: coordinate maint operations through lock owner (Jason Dillaman)
- librbd: copy-on-read (Min Chen, Li Wang, Yunchuan Wen, Cheng Cheng, Jason Dillaman)
- librbd: enforce write ordering with a snapshot (Jason Dillaman)
- librbd: fadvise-style hints; add misc hints for certain operations (Jianpeng Ma)
- librbd: fix coverity false-positives (Jason Dillaman)
- librbd: fix snap create races (Jason Dillaman)
- librbd: flush AIO operations asynchronously (#10714 Jason Dillaman)
- librbd: make async versions of long-running maint operations (Jason Dillaman)
- librbd: mock tests (Jason Dillaman)
- librbd: optionally blacklist clients before breaking locks (#10761 Jason Dillaman)
- librbd: prevent copyup during shrink (Jason Dillaman)
- mds: add cephfs-table-tool (John Spray)
- mds: avoid sending traceless replies in most cases (Yan, Zheng)
- mds: export dir asok command (John Spray)
- mds: fix stray/purge perfcounters (#10388 John Spray)
- mds: handle heartbeat\_reset during shutdown (#10382 John Spray)
- mds: many many snapshot-related fixes (Yan, Zheng)
- mds: refactor, improve Session storage (John Spray)

- misc coverity fixes (Danny Al-Gaaf)
- mon: add noforward flag for some mon commands (Mykola Golub)
- mon: disallow empty pool names (#10555 Wido den Hollander)
- mon: do not deactivate last mds (#10862 John Spray)
- mon: drop old ceph\_mon\_store\_converter (Sage Weil)
- mon: fix 'ceph pg dump\_stuck degraded' (Xinxin Shu)
- mon: fix 'profile osd' use of config-key function on mon (#10844 Joao Eduardo Luis)
- mon: fix compatset initialization during mkfs (Joao Eduardo Luis)
- mon: fix feature tracking during elections (Joao Eduardo Luis)
- mon: fix mds gid/rank/state parsing (John Spray)
- mon: ignore failure reports from before up\_from (#10762 Dan van der Ster, Sage Weil)
- mon: improved corrupt CRUSH map detection (Joao Eduardo Luis)
- mon: include pg\_temp count in osdmap summary (Sage Weil)
- mon: log health summary to cluster log (#9440 Joao Eduardo Luis)
- mon: make 'mds fail' idempotent (John Spray)
- mon: make pg dump {sum,pgs,pgs\_brief} work for format=plain (#5963 #6759 Mykola Golub)
- mon: new pool safety flags nodelete, nopgchange, nosizechange (#9792 Mykola Golub)
- mon: new, friendly 'ceph pg ls ...' command (Xinxin Shu)
- mon: prevent MDS transition from STOPPING (#10791 Greg Farnum)
- mon: propose all pending work in one transaction (Sage Weil)
- mon: remove pg\_temps for nonexistent pools (Joao Eduardo Luis)
- mon: require mon\_allow\_pool\_delete option to remove pools (Sage Weil)
- mon: set globalid prealloc to larger value (Sage Weil)
- mon: skip zeroed osd stats in get\_rule\_avail (#10257 Joao Eduardo Luis)
- mon: validate min\_size range (Jianpeng Ma)

- msgr: async: bind threads to CPU cores, improved poll (Haomai Wang)
- msgr: fix crc configuration (Mykola Golub)
- msgr: misc unit tests (Haomai Wang)
- msgr: xio: XioMessenger RDMA support (Casey Bodley, Vu Pham, Matt Benjamin)
- osd, librados: fadvise-style librados hints (Jianpeng Ma)
- osd, librados: fix xattr\_cmp\_u64 (Dongmao Zhang)
- osd,mon: add 'norebalance' flag (Kefu Chai)
- osd,mon: specify OSD features explicitly in MOSDBoot (#10911 Sage Weil)
- osd: add option to prioritize heartbeat network traffic (Jian Wen)
- osd: add support for the SHEC erasure-code algorithm (Takeshi Miyamae, Loic Dachary)
- osd: allow recovery while below min\_size (Samuel Just)
- osd: allow recovery with fewer than min\_size OSDs (Samuel Just)
- osd: allow writes to degraded objects (Samuel Just)
- osd: allow writes to degraded objects (Samuel Just)
- osd: avoid publishing unchanged PG stats (Sage Weil)
- osd: cache recent ObjectContexts (Dong Yuan)
- osd: clear cache on interval change (Samuel Just)
- osd: do no proxy reads unless target OSDs are new (#10788 Sage Weil)
- osd: do not update digest on inconsistent object (#10524 Samuel Just)
- osd: don't record digests for snapdirs (#10536 Samuel Just)
- osd: fix OSDCap parser on old (el6) boost::spirit (#10757 Kefu Chai)
- osd: fix OSDCap parsing on el6 (#10757 Kefu Chai)
- osd: fix ObjectStore::Transaction encoding version (#10734 Samuel Just)
- osd: fix auth object selection during repair (#10524 Samuel Just)
- osd: fix bug in pending digest updates (#10840 Samuel Just)
- osd: fix cancel\_proxy\_read\_ops (Sage Weil)
- osd: fix cleanup of interrupted pg deletion (#10617 Sage Weil)

- osd: fix journal wrapping bug (#10883 David Zafman)
- osd: fix leak in SnapTrimWQ (#10421 Kefu Chai)
- osd: fix memstore free space calculation (Xiaoxi Chen)
- osd: fix mixed-version peering issues (Samuel Just)
- osd: fix object digest update bug (#10840 Samuel Just)
- osd: fix ordering issue with new transaction encoding (#10534 Dong Yuan)
- osd: fix past\_interval generation (#10427 #10430 David Zafman)
- osd: fix short read handling on push (#8121 David Zafman)
- osd: fix watch timeout cache state update (#10784 David Zafman)
- osd: force promotion of watch/notify ops (Zhiqiang Wang)
- osd: improve idempotency detection across cache promotion/demotion (#8935 Sage Weil, Samuel Just)
- osd: include activating peers in blocked\_by (#10477 Sage Weil)
- osd: jerasure and gf-complete updates from upstream (#10216 Loic Dachary)
- osd: journal: check fsync/fdatasync result (Jianpeng Ma)
- osd: journal: fix hang on shutdown (#10474 David Zafman)
- osd: journal: fix header.committed\_up\_to (Xinze Chi)
- osd: journal: initialize throttle (Ning Yao)
- osd: journal: misc bug fixes (#6003 David Zafman, Samuel Just)
- osd: misc cleanup (Xinze Chi, Yongyue Sun)
- osd: new 'activating' state between peering and active (Sage Weil)
- osd: preserve reqids for idempotency checks for promote/demote (Sage Weil, Zhiqiang Wang, Samuel Just)
- osd: remove dead locking code (Xinxin Shu)
- osd: restrict scrub to certain times of day (Xinze Chi)
- osd: rocksdb: fix shutdown (Hoamai Wang)
- pybind: fix memory leak in librados bindings (Billy Olsen)
- qa: fix mds tests (#10539 John Spray)

- qa: ignore duplicates in rados ls (Josh Durgin)
- qa: improve hadoop tests (Noah Watkins)
- qa: reorg fs quota tests (Greg Farnum)
- rados: fix usage (Jianpeng Ma)
- radosgw-admin: add replicalog update command (Yehuda Sadeh)
- rbd-fuse: clean up on shutdown (Josh Durgin)
- rbd: add ‘merge-diff’ function (MingXin Liu, Yunchuan Wen, Li Wang)
- rbd: fix buffer handling on image import (#10590 Jason Dillaman)
- rbd: leave exclusive lockin goff by default (Jason Dillaman)
- rbd: update init-rbdmap to fix dup mount point (Karel Striegel)
- rbd: use IO hints for import, export, and bench operations (#10462 Jason Dillaman)
- rbd\_recover\_tool: RBD image recovery tool (Min Chen)
- rgw: S3-style object versioning support (Yehuda Sadeh)
- rgw: check keystone auth for S3 POST requests (#10062 Abhishek Lekshmanan)
- rgw: extend replica log API (purge-all) (Yehuda Sadeh)
- rgw: fail S3 POST if keystone not configured (#10688 Valery Tschopp, Yehuda Sadeh)
- rgw: fix XML header on get ACL request (#10106 Yehuda Sadeh)
- rgw: fix bucket removal with data purge (Yehuda Sadeh)
- rgw: fix replica log indexing (#8251 Yehuda Sadeh)
- rgw: fix swift metadata header name (Dmytro Iurchenko)
- rgw: remove multipart entries from bucket index on abort (#10719 Yehuda Sadeh)
- rgw: respond with 204 to POST on containers (#10667 Yuan Zhou)
- rgw: reuse fcgx connection struct (#10194 Yehuda Sadeh)
- rgw: support multiple host names (#7467 Yehuda Sadeh)
- rgw: swift: dump container’s custom metadata (#10665 Ahmad Faheem, Dmytro Iurchenko)
- rgw: swift: support Accept header for response format (#10746 Dmytro Iurchenko)

- rgw: swift: support for X-Remove-Container-Meta-{key} (#10475 Dmytro Iurchenko)
- rpm: move rgw logrotate to radosgw subpackage (Ken Dreyer)
- tests: centos-6 dockerfile (#10755 Loic Dachary)
- tests: unit tests for shared\_cache (Dong Yuan)
- vstart.sh: work with cmake (Yehuda Sadeh)

## v0.92

---

This is the second-to-last chunk of new stuff before Hammer. Big items include additional checksums on OSD objects, proxied reads in the cache tier, image locking in RBD, optimized OSD Transaction and replication messages, and a big pile of RGW and MDS bug fixes.

## Upgrading

---

- The experimental ‘keyvaluestore-dev’ OSD backend has been renamed ‘keyvaluestore’ (for simplicity) and marked as experimental. To enable this untested feature and acknowledge that you understand that it is untested and may destroy data, you need to add the following to your ceph.conf:

```
1. enable experimental unrecoverable data corrupting features = keyvaluestore
```

- The following librados C API function calls take a ‘flags’ argument whose value is now correctly interpreted:

```
rados_write_op_operate() rados_aio_write_op_operate() rados_read_op_operate() rados_aio_read_op_operate()
```

The flags were not correctly being translated from the librados constants to the internal values. Now they are. Any code that is passing flags to these methods should be audited to ensure that they are using the correct LIBRADOS\_OP\_FLAG\_\* constants.

- The ‘rados’ CLI ‘copy’ and ‘cppool’ commands now use the copy-from operation, which means the latest CLI cannot run these commands against pre-firefly OSDs.
- The librados watch/notify API now includes a watch\_flush() operation to flush the async queue of notify operations. This should be called by any watch/notify user prior to rados\_shutdown().

## Notable Changes

---

- add experimental features option (Sage Weil)

- build: fix 'make check' races (#10384 Loic Dachary)
- build: fix pkg names when libkeyutils is missing (Pankag Garg, Ken Dreyer)
- ceph: make 'ceph -s' show PG state counts in sorted order (Sage Weil)
- ceph: make 'ceph tell mon.\* version' work (Mykola Golub)
- ceph-monstore-tool: fix/improve CLI (Joao Eduardo Luis)
- ceph: show primary-affinity in 'ceph osd tree' (Mykola Golub)
- common: add TableFormatter (Andreas Peters)
- common: check syncfs() return code (Jianpeng Ma)
- doc: do not suggest dangerous XFS nobarrier option (Dan van der Ster)
- doc: misc updates (Nilamdyuti Goswami, John Wilkins)
- install-deps.sh: do not require sudo when root (Loic Dachary)
- libcephfs: fix dirfrag trimming (#10387 Yan, Zheng)
- libcephfs: fix mount timeout (#10041 Yan, Zheng)
- libcephfs: fix test (#10415 Yan, Zheng)
- libcephfs: fix use-afer-free on umount (#10412 Yan, Zheng)
- libcephfs: include ceph and git version in client metadata (Sage Weil)
- librados: add watch\_flush() operation (Sage Weil, Haomai Wang)
- librados: avoid memcpy on getxattr, read (Jianpeng Ma)
- librados: create ioctx by pool id (Jason Dillaman)
- librados: do notify completion in fast-dispatch (Sage Weil)
- librados: remove shadowed variable (Kefu Chain)
- librados: translate op flags from C APIs (Matthew Richards)
- librbd: differentiate between R/O vs R/W features (Jason Dillaman)
- librbd: exclusive image locking (Jason Dillaman)
- librbd: fix write vs import race (#10590 Jason Dillaman)
- librbd: gracefully handle deleted/renamed pools (#10270 Jason Dillaman)
- mds: asok command for fetching subtree map (John Spray)

- mds: constify MDSCacheObjects (John Spray)
- misc: various valgrind fixes and cleanups (Danny Al-Gaaf)
- mon: fix 'mds fail' for standby MDSs (John Spray)
- mon: fix stashed monmap encoding (#5203 Xie Rui)
- mon: implement 'fs reset' command (John Spray)
- mon: respect down flag when promoting standbys (John Spray)
- mount.ceph: fix suprious error message (#10351 Yan, Zheng)
- msgr: async: many fixes, unit tests (Haomai Wang)
- msgr: simple: retry binding to port on failure (#10029 Wido den Hollander)
- osd: add fadvise flags to ObjectStore API (Jianpeng Ma)
- osd: add get\_latest\_osdmap asok command (#9483 #9484 Mykola Golub)
- osd: EIO on whole-object reads when checksum is wrong (Sage Weil)
- osd: filejournal: don't cache journal when not using direct IO (Jianpeng Ma)
- osd: fix ioprio option (Mykola Golub)
- osd: fix scrub delay bug (#10693 Samuel Just)
- osd: fix watch reconnect race (#10441 Sage Weil)
- osd: handle no-op write with snapshot (#10262 Sage Weil)
- osd: journal: fix journal zeroing when direct IO is enabled (Xie Rui)
- osd: keyvaluestore: cleanup dead code (Ning Yao)
- osd, mds: 'ops' as shorthand for 'dump\_ops\_in\_flight' on asok (Sage Weil)
- osd: memstore: fix size limit (Xiaoxi Chen)
- osd: misc scrub fixes (#10017 Loic Dachary)
- osd: new optimized encoding for ObjectStore::Transaction (Dong Yuan)
- osd: optimize filter\_snapc (Ning Yao)
- osd: optimize WBThrottle map with unordered\_map (Ning Yao)
- osd: proxy reads during cache promote (Zhiqiang Wang)
- osd: proxy read support (Zhiqiang Wang)

- osd: remove legacy classic scrub code (Sage Weil)
- osd: remove unused fields in MOSDSubOp (Xiaoxi Chen)
- osd: replace MOSDSubOp messages with simpler, optimized MOSDRepOp (Xiaoxi Chen)
- osd: store whole-object checksums on scrub, write\_full (Sage Weil)
- osd: verify kernel is new enough before using XFS extsize ioctl, enable by default (#9956 Sage Weil)
- rados: use copy-from operation for copy, cppool (Sage Weil)
- rgw: change multipart upload id magic (#10271 Yehuda Sadeh)
- rgw: decode http query params correction (#10271 Yehuda Sadeh)
- rgw: fix content length check (#10701 Axel Dunkel, Yehuda Sadeh)
- rgw: fix partial GET in swift (#10553 Yehuda Sadeh)
- rgw: fix shutdown (#10472 Yehuda Sadeh)
- rgw: include XML ns on get ACL request (#10106 Yehuda Sadeh)
- rgw: misc fixes (#10307 Yehuda Sadeh)
- rgw: only track cleanup for objects we write (#10311 Yehuda Sadeh)
- rgw: tweak error codes (#10329 #10334 Yehuda Sadeh)
- rgw: use gc for multipart abort (#10445 Aaron Bassett, Yehuda Sadeh)
- sysvinit: fix race in 'stop' (#10389 Loic Dachary)
- test: fix bufferlist tests (Jianpeng Ma)
- tests: improve docker-based tests (Loic Dachary)

## v0.91

---

We are quickly approaching the Hammer feature freeze but have a few more dev releases to go before we get there. The headline items are subtree-based quota support in CephFS (ceph-fuse/libcephfs client support only for now), a rewrite of the watch/notify librados API used by RBD and RGW, OSDMap checksums to ensure that maps are always consistent inside the cluster, new API calls in librados and librbd for IO hinting modeled after posix\_fadvise, and improved storage of per-PG state.

We expect two more releases before the Hammer feature freeze (v0.93).

## Upgrading

- The ‘category’ field for objects has been removed. This was originally added to track PG stat summations over different categories of objects for use by radosgw. It is no longer has any known users and is prone to abuse because it can lead to a pg\_stat\_t structure that is unbounded. The librados API calls that accept this field now ignore it, and the OSD no longer tracks the per-category summations.
- The output for ‘rados df’ has changed. The ‘category’ level has been eliminated, so there is now a single stat object per pool. The structure of the JSON output is different, and the plaintext output has one less column.
- The ‘rados create <objectname> [category]’ optional category argument is no longer supported or recognized.
- rados.py’s Rados class no longer has a `__del__` method; it was causing problems on interpreter shutdown and use of threads. If your code has Rados objects with limited lifetimes and you’re concerned about locked resources, call `Rados.shutdown()` explicitly.
- There is a new version of the librados watch/notify API with vastly improved semantics. Any applications using this interface are encouraged to migrate to the new API. The old API calls are marked as deprecated and will eventually be removed.
- The librados `rados_unwatch()` call used to be safe to call on an invalid handle. The new version has undefined behavior when passed a bogus value (for example, when `rados_watch()` returns an error and handle is not defined).
- The structure of the formatted ‘pg stat’ command is changed for the portion that counts states by name to avoid using the ‘+’ character (which appears in state names) as part of the XML token (it is not legal).

## Notable Changes

- `asyncmsgc: misc fixes` (Haomai Wang)
- `buffer: add ‘shareable’ construct` (Matt Benjamin)
- `build: aarch64 build fixes` (Noah Watkins, Haomai Wang)
- `build: support for jemalloc` (Shishir Gowda)
- `ceph-disk: allow journal partition re-use` (#10146 Loic Dachary, Dav van der Ster)
- `ceph-disk: misc fixes` (Christos Stavrakakis)
- `ceph-fuse: fix kernel cache trimming` (#10277 Yan, Zheng)
- `ceph-objectstore-tool: many many improvements` (David Zafman)

- common: support new gperftools header locations (Key Dreyer)
- crush: straw bucket weight calculation fixes (#9998 Sage Weil)
- doc: misc improvements (Nilamdyuti Goswami, John Wilkins, Chris Holcombe)
- libcephfs, ceph-fuse: add ‘status’ asok (John Spray)
- librados, osd: new watch/notify implementation (Sage Weil)
- librados: drop ‘category’ feature (Sage Weil)
- librados: fix pool deletion handling (#10372 Sage Weil)
- librados: new fadvise API (Ma Jianpeng)
- libradosstriper: fix remove() (Dongmao Zhang)
- librbd: complete pending ops before closing image (#10299 Josh Durgin)
- librbd: fadvise API (Ma Jianpeng)
- mds: ENOSPC and OSDMap epoch barriers (#7317 John Spray)
- mds: dirfrag buf fix (Yan, Zheng)
- mds: disallow most commands on inactive MDS’s (Greg Farnum)
- mds: drop dentries, leases on deleted directories (#10164 Yan, Zheng)
- mds: handle zero-size xattr (#10335 Yan, Zheng)
- mds: subtree quota support (Yunchuan Wen)
- memstore: free space tracking (John Spray)
- misc cleanup (Danny Al-Gaaf, David Anderson)
- mon: ‘osd crush reweight-all’ command (Sage Weil)
- mon: allow full flag to be manually cleared (#9323 Sage Weil)
- mon: delay failure injection (Joao Eduardo Luis)
- mon: fix paxos timeouts (#10220 Joao Eduardo Luis)
- mon: get canonical OSDMap from leader (#10422 Sage Weil)
- msgr: fix RESETSESSION bug (#10080 Greg Farnum)
- objectstore: deprecate collection attrs (Sage Weil)
- osd, mon: add checksums to all OSDMaps (Sage Weil)

- osd: allow deletion of objects with watcher (#2339 Sage Weil)
- osd: allow sparse read for Push/Pull (Haomai Wang)
- osd: cache reverse\_nibbles hash value (Dong Yuan)
- osd: drop upgrade support for pre-dumpling (Sage Weil)
- osd: enable and use posix\_fadvise (Sage Weil)
- osd: erasure-code: enforce chunk size alignment (#10211 Loic Dachary)
- osd: erasure-code: jerasure support for NEON (Loic Dachary)
- osd: erasure-code: relax cauchy w restrictions (#10325 David Zhang, Loic Dachary)
- osd: erasure-code: update gf-complete to latest upstream (Loic Dachary)
- osd: fix WBTHrottle perf counters (Haomai Wang)
- osd: fix backfill bug (#10150 Samuel Just)
- osd: fix occasional peering stalls (#10431 Sage Weil)
- osd: fix scrub vs try-flush bug (#8011 Samuel Just)
- osd: fix stderr with -f or -d (Dan Mick)
- osd: misc FIEMAP fixes (Ma Jianpeng)
- osd: optimize Finisher (Xinze Chi)
- osd: store PG metadata in per-collection objects for better concurrency (Sage Weil)
- pyrados: add object lock support (#6114 Mehdi Abaakouk)
- pyrados: fix misnamed wait\_\* routings (#10104 Dan Mick)
- pyrados: misc cleanups (Kefu Chai)
- qa: add large auth ticket tests (Ilya Dryomov)
- qa: many ‘make check’ improvements (Loic Dachary)
- qa: misc tests (Loic Dachary, Yan, Zheng)
- rgw: conditional PUT on ETag (#8562 Ray Lv)
- rgw: fix error codes (#10334 #10329 Yehuda Sadeh)
- rgw: index swift keys appropriately (#10471 Yehuda Sadeh)
- rgw: prevent illegal bucket policy that doesn’t match placement rule (Yehuda Sadeh)

Sadeh)

- rgw: run radosgw as apache with systemd (#10125 Loic Dachary)
- rgw: support X-Storage-Policy header for Swift storage policy compat (Yehuda Sadeh)
- rgw: use rn for http headers (#9254 Yehuda Sadeh)
- rpm: misc fixes (Key Dreyer)

## v0.90

---

This is the last development release before Christmas. There are some API cleanups for librados and librbd, and lots of bug fixes across the board for the OSD, MDS, RGW, and CRUSH. The OSD also gets support for discard (potentially helpful on SSDs, although it is off by default), and there are several improvements to ceph-disk.

The next two development releases will be getting a slew of new functionality for hammer. Stay tuned!

## Upgrading

---

- Previously, the formatted output of ‘ceph pg stat -f ...’ was a full pg dump that included all metadata about all PGs in the system. It is now a concise summary of high-level PG stats, just like the unformatted ‘ceph pg stat’ command.
- All JSON dumps of floating point values were incorrectly surrounding the value with quotes. These quotes have been removed. Any consumer of structured JSON output that was consuming the floating point values was previously having to interpret the quoted string and will most likely need to be fixed to take the unquoted number.

## Notable Changes

---

- arch: fix NEON feaeture detection (#10185 Loic Dachary)
- build: adjust build deps for yasm, virtualenv (Jianpeng Ma)
- build: improve build dependency tooling (Loic Dachary)
- ceph-disk: call partx/partprobe consistency (#9721 Loic Dachary)
- ceph-disk: fix dmcrypt key permissions (Loic Dachary)
- ceph-disk: fix umount race condition (#10096 Blaine Gardner)
- ceph-disk: init=none option (Loic Dachary)

- ceph-monstore-tool: fix shutdown (#10093 Loic Dachary)
- ceph-objectstore-tool: fix import (#10090 David Zafman)
- ceph-objectstore-tool: many improvements and tests (David Zafman)
- ceph.spec: package rbd-replay-prep (Ken Dreyer)
- common: add ‘perf reset ...’ admin command (Jianpeng Ma)
- common: do not unlock rwlock on destruction (Federico Simoncelli)
- common: fix block device discard check (#10296 Sage Weil)
- common: remove broken CEPH\_LOCKDEP optoin (Kefu Chai)
- crush: fix tree bucket behavior (Rongze Zhu)
- doc: add build-doc guidlines for Fedora and CentOS/RHEL (Nilamdyuti Goswami)
- doc: enable rbd cache on openstack deployments (Sebastien Han)
- doc: improved installation nots on CentOS/RHEL installs (John Wilkins)
- doc: misc cleanups (Adam Spiers, Sebastien Han, Nilamdyuti Goswami, Ken Dreyer, John Wilkins)
- doc: new man pages (Nilamdyuti Goswami)
- doc: update release descriptions (Ken Dreyer)
- doc: update sepia hardware inventory (Sandon Van Ness)
- librados: only export public API symbols (Jason Dillaman)
- libradosstriper: fix stat strtoll (Dongmao Zhang)
- libradosstriper: fix trunc method (#10129 Sebastien Ponce)
- librbd: fix list\_children from invalid pool ioctxs (#10123 Jason Dillaman)
- librbd: only export public API symbols (Jason Dillaman)
- many coverity fixes (Danny Al-Gaaf)
- mds: ‘flush journal’ admin command (John Spray)
- mds: fix MDLog IO callback deadlock (John Spray)
- mds: fix deadlock during journal probe vs purge (#10229 Yan, Zheng)
- mds: fix race trimming log segments (Yan, Zheng)
- mds: store backtrace for stray dir (Yan, Zheng)

- mds: verify backtrace when fetching dirfrag (#9557 Yan, Zheng)
- mon: add max pgs per osd warning (Sage Weil)
- mon: fix \_ratio units and types (Sage Weil)
- mon: fix JSON dumps to dump floats as floats and not strings (Sage Weil)
- mon: fix formatter 'pg stat' command output (Sage Weil)
- msgr: async: several fixes (Haomai Wang)
- msgr: simple: fix rare deadlock (Greg Farnum)
- osd: batch pg log trim (Xinze Chi)
- osd: clean up internal ObjectStore interface (Sage Weil)
- osd: do not abort deep scrub on missing hinfo (#10018 Loic Dachary)
- osd: fix ghobject\_t formatted output to include shard (#10063 Loic Dachary)
- osd: fix osd peer check on scrub messages (#9555 Sage Weil)
- osd: fix pgls filter ops (#9439 David Zafman)
- osd: flush snapshots from cache tier immediately (Sage Weil)
- osd: keyvaluestore: fix getattr semantics (Haomai Wang)
- osd: keyvaluestore: fix key ordering (#10119 Haomai Wang)
- osd: limit in-flight read requests (Jason Dillaman)
- osd: log when scrub or repair starts (Loic Dachary)
- osd: support for discard for journal trim (Jianpeng Ma)
- qa: fix osd create dup tests (#10083 Loic Dachary)
- rgw: add location header when object is in another region (VRan Liu)
- rgw: check timestamp on s3 keystone auth (#10062 Abhishek Lekshmanan)
- rgw: make sysvinit script set ulimit -n properly (Sage Weil)
- systemd: better systemd unit files (Owen Syngue)
- tests: ability to run unit tests under docker (Loic Dachary)

## v0.89

This is the second development release since Giant. The big items include the first

batch of scrub patches from Greg for CephFS, a rework in the librados object listing API to properly handle namespaces, and a pile of bug fixes for RGW. There are also several smaller issues fixed up in the performance area with buffer alignment and memory copies, osd cache tiering agent, and various CephFS fixes.

## Upgrading

---

- New ability to list all objects from all namespaces can fail or return incomplete results when not all OSDs have been upgraded. Features rados -all ls, rados cpool, rados export, rados cache-flush-evict-all and rados cache-try-flush-evict-all can also fail or return incomplete results.

## Notable Changes

---

- buffer: add list::get\_contiguous (Sage Weil)
- buffer: avoid rebuild if buffer already contiguous (Jianpeng Ma)
- ceph-disk: improved systemd support (Owen Synge)
- ceph-disk: set guid if reusing journal partition (Dan van der Ster)
- ceph-fuse, libcephfs: allow xattr caps in inject\_release\_failure (#9800 John Spray)
- ceph-fuse, libcephfs: fix I\_COMPLETE\_ORDERED checks (#9894 Yan, Zheng)
- ceph-fuse: fix dentry invalidation on 3.18+ kernels (#9997 Yan, Zheng)
- crush: fix detach\_bucket (#10095 Sage Weil)
- crush: fix several bugs in adjust\_item\_weight (Rongze Zhu)
- doc: add dumpling to firefly upgrade section (#7679 John Wilkins)
- doc: document erasure coded pool operations (#9970 Loic Dachary)
- doc: file system osd config settings (Kevin Dalley)
- doc: key/value store config reference (John Wilkins)
- doc: update openstack docs for Juno (Sebastien Han)
- fix cluster logging from non-mon daemons (Sage Weil)
- init-ceph: check for systemd-run before using it (Boris Ranto)
- librados: fix infinite loop with skipped map epochs (#9986 Ding Dinghua)
- librados: fix iterator operator= bugs (#10082 David Zafman, Yehuda Sadeh)

- librados: fix null deref when pool DNE (#9944 Sage Weil)
- librados: fix timer race from recent refactor (Sage Weil)
- libradosstriper: fix shutdown hang (Dongmao Zhang)
- librbd: don't close a closed parent in failure path (#10030 Jason Dillaman)
- librbd: fix diff test (#10002 Josh Durgin)
- librbd: fix locking for readahead (#10045 Jason Dillaman)
- librbd: refactor unit tests to use fixtures (Jason Dillaman)
- many many coverity cleanups (Danny Al-Gaaf)
- mds: a whole bunch of initial scrub infrastructure (Greg Farnum)
- mds: fix compat\_version for MClientSession (#9945 John Spray)
- mds: fix reply snapbl (Yan, Zheng)
- mon: allow adding tiers to fs pools (#10135 John Spray)
- mon: fix MDS health status from peons (#10151 John Spray)
- mon: fix caching for min\_last\_epoch\_clean (#9987 Sage Weil)
- mon: fix error output for add\_data\_pool (#9852 Joao Eduardo Luis)
- mon: include entity name in audit log for forwarded requests (#9913 Joao Eduardo Luis)
- mon: paxos: allow reads while proposing (#9321 #9322 Joao Eduardo Luis)
- msgr: asyncmessenger: add kqueue support (#9926 Haomai Wang)
- osd, librados: revamp PG listing API to handle namespaces (#9031 #9262 #9438 David Zafman)
- osd, mon: send intial pg create time from mon to osd (#9887 David Zafman)
- osd: allow whiteout deletion in cache pool (Sage Weil)
- osd: cache pool: ignore min flush age when cache is full (Xinze Chi)
- osd: erasure coding: allow bench.sh to test ISA backend (Yuan Zhou)
- osd: erasure-code: encoding regression tests, corpus (#9420 Loic Dachary)
- osd: fix journal shutdown race (Sage Weil)
- osd: fix object age eviction (Zhiqiang Wang)

- osd: fix object atime calculation (Xinze Chi)
- osd: fix past\_interval display bug (#9752 Loic Dachary)
- osd: journal: fix alignment checks, avoid useless memmove (Jianpeng Ma)
- osd: journal: update committed\_thru after replay (#6756 Samuel Just)
- osd: keyvaluestore\_dev: optimization (Chendi Xue)
- osd: make misdirected op checks robust for EC pools (#9835 Sage Weil)
- osd: removed some dead code (Xinze Chi)
- qa: parallelize make check (Loic Dachary)
- qa: tolerate nearly-full disk for make check (Loic Dachary)
- rgw: create subuser if needed when creating user (#10103 Yehuda Sadeh)
- rgw: fix If-Modified-Since (VRan Liu)
- rgw: fix content-length update (#9576 Yehuda Sadeh)
- rgw: fix disabling of max\_size quota (#9907 Dong Lei)
- rgw: fix incorrect len when len is 0 (#9877 Yehuda Sadeh)
- rgw: fix object copy content type (#9478 Yehuda Sadeh)
- rgw: fix user stags in get-user-info API (#9359 Ray Lv)
- rgw: remove swift user manifest (DLO) hash calculation (#9973 Yehuda Sadeh)
- rgw: return timestamp on GET/HEAD (#8911 Yehuda Sadeh)
- rgw: set ETag on object copy (#9479 Yehuda Sadeh)
- rgw: update bucket index on attr changes, for multi-site sync (#5595 Yehuda Sadeh)

## v0.88

---

This is the first development release after Giant. The two main features merged this round are the new AsyncMessenger (an alternative implementation of the network layer) from Haomai Wang at UnitedStack, and support for POSIX file locks in ceph-fuse and libcephfs from Yan, Zheng. There is also a big pile of smaller items that were merged while we were stabilizing Giant, including a range of smaller performance and bug fixes and some new tracepoints for LTTNG.

## Notable Changes

- ceph-disk: Scientific Linux support (Dan van der Ster)
- ceph-disk: respect -statedir for keyring (Loic Dachary)
- ceph-fuse, libcephfs: POSIX file lock support (Yan, Zheng)
- ceph-fuse, libcephfs: fix cap flush overflow (Greg Farnum, Yan, Zheng)
- ceph-fuse, libcephfs: fix root inode xattrs (Yan, Zheng)
- ceph-fuse, libcephfs: preserve dir ordering (#9178 Yan, Zheng)
- ceph-fuse, libcephfs: trim inodes before reconnecting to MDS (Yan, Zheng)
- ceph: do not parse injectargs twice (Loic Dachary)
- ceph: make 'ceph -s' output more readable (Sage Weil)
- ceph: new 'ceph tell mds.\$name\_or\_rank\_or\_gid' (John Spray)
- ceph: test robustness (Joao Eduardo Luis)
- ceph\_objectstore\_tool: behave with sharded flag (#9661 David Zafman)
- cephfs-journal-tool: fix journal import (#10025 John Spray)
- cephfs-journal-tool: skip up to expire\_pos (#9977 John Spray)
- cleanup rados.h definitions with macros (Ilya Dryomov)
- common: shared\_cache unit tests (Cheng Cheng)
- config: add \$cctid meta variable (Adam Crume)
- crush: fix buffer overrun for poorly formed rules (#9492 Johnu George)
- crush: improve constness (Loic Dachary)
- crushtool: add -location <id> command (Sage Weil, Loic Dachary)
- default to libnss instead of crypto++ (Federico Gimenez)
- doc: ceph osd reweight vs crush weight (Laurent Guerby)
- doc: document the LRC per-layer plugin configuration (Yuan Zhou)
- doc: erasure code doc updates (Loic Dachary)
- doc: misc updates (Alfredo Deza, VRan Liu)
- doc: preflight doc fixes (John Wilkins)
- doc: update PG count guide (Gerben Meijer, Laurent Guerby, Loic Dachary)

- keyvaluestore: misc fixes (Haomai Wang)
- keyvaluestore: performance improvements (Haomai Wang)
- librados: add rados\_pool\_get\_base\_tier() call (Adam Crume)
- librados: cap buffer length (Loic Dachary)
- librados: fix objecter races (#9617 Josh Durgin)
- libradosstriper: misc fixes (Sebastien Ponce)
- librbd: add missing python docstrings (Jason Dillaman)
- librbd: add readahead (Adam Crume)
- librbd: fix cache tiers in list\_children and snap\_unprotect (Adam Crume)
- librbd: fix performance regression in ObjectCacher (#9513 Adam Crume)
- librbd: lttng tracepoints (Adam Crume)
- librbd: misc fixes (Xinxin Shu, Jason Dillaman)
- mds: fix sessionmap lifecycle bugs (Yan, Zheng)
- mds: initialize root inode xattr version (Yan, Zheng)
- mds: introduce auth caps (John Spray)
- mds: misc bugs (Greg Farnum, John Spray, Yan, Zheng, Henry Change)
- misc coverity fixes (Danny Al-Gaaf)
- mon: add 'ceph osd rename-bucket ...' command (Loic Dachary)
- mon: clean up auth list output (Loic Dachary)
- mon: fix 'osd crush link' id resolution (John Spray)
- mon: fix misc error paths (Joao Eduardo Luis)
- mon: fix paxos off-by-one corner case (#9301 Sage Weil)
- mon: new 'ceph pool ls [detail]' command (Sage Weil)
- mon: wait for writeable before cross-proposing (#9794 Joao Eduardo Luis)
- msgr: avoid useless new/delete (Haomai Wang)
- msgr: fix delay injection bug (#9910 Sage Weil, Greg Farnum)
- msgr: new AsymcMessenger alternative implementation (Haomai Wang)

- msgr: prefetch data when doing recv (Yehuda Sadeh)
- osd: add erasure code corpus (Loic Dachary)
- osd: add misc tests (Loic Dachary, Danny Al-Gaaf)
- osd: cleanup boost optionals (William Kennington)
- osd: expose non-journal backends via ceph-osd CLI (Hoamai Wang)
- osd: fix JSON output for stray OSDs (Loic Dachary)
- osd: fix ioprio options (Loic Dachary)
- osd: fix transaction accounting (Jianpeng Ma)
- osd: misc optimizations (Xinxin Shu, Zhiqiang Wang, Xinze Chi)
- osd: use FIEMAP\_FLAGS\_SYNC instead of fsync (Jianpeng Ma)
- rados: fix put of /dev/null (Loic Dachary)
- rados: parse command-line arguments more strictly (#8983 Adam Crume)
- rbd-fuse: fix memory leak (Adam Crume)
- rbd-replay-many (Adam Crume)
- rbd-replay: --anonymize flag to rbd-replay-prep (Adam Crume)
- rbd: fix 'rbd diff' for non-existent objects (Adam Crume)
- rbd: fix error when striping with format 1 (Sebastien Han)
- rbd: fix export for image sizes over 2GB (Vicente Cheng)
- rbd: use rolling average for rbd bench-write throughput (Jason Dillaman)
- rgw: send explicit HTTP status string (Yehuda Sadeh)
- rgw: set length for keystone token validation request (#7796 Yehuda Sadeh, Mark Kirkwood)
- udev: fix rules for CentOS7/RHEL7 (Loic Dachary)
- use clock\_gettime instead of gettimeofday (Jianpeng Ma)
- vstart.sh: set up environment for s3-tests (Luis Pabon)

# v0.87.2 Giant

This is the second (and possibly final) point release for Giant.

We recommend all v0.87.x Giant users upgrade to this release.

## Notable Changes

- ceph-objectstore-tool: only output unsupported features when incompatible (#11176 David Zafman)
- common: do not implicitly unlock rwlock on destruction (Federico Simoncelli)
- common: make wait timeout on empty queue configurable (#10818 Samuel Just)
- crush: pick ruleset id that matches and rule id (Xiaoxi Chen)
- crush: set\_choose\_tries = 100 for new erasure code rulesets (#10353 Loic Dachary)
- librados: check initialized atomic safely (#9617 Josh Durgin)
- librados: fix failed tick\_event assert (#11183 Zhiqiang Wang)
- librados: fix looping on skipped maps (#9986 Ding Dinghua)
- librados: fix op submit with timeout (#10340 Samuel Just)
- librados: pybind: fix memory leak (#10723 Billy Olsen)
- librados: pybind: keep reference to callbacks (#10775 Josh Durgin)
- librados: translate operation flags from C APIs (Matthew Richards)
- libradosstriper: fix write\_full on ENOENT (#10758 Sebastien Ponce)
- libradosstriper: use strtoll instead of strtol (Dongmao Zhang)
- mds: fix assertion caused by system time moving backwards (#11053 Yan, Zheng)
- mon: allow injection of random delays on writes (Joao Eduardo Luis)
- mon: do not trust small osd epoch cache values (#10787 Sage Weil)
- mon: fail non-blocking flush if object is being scrubbed (#8011 Samuel Just)
- mon: fix division by zero in stats dump (Joao Eduardo Luis)
- mon: fix get\_rule\_avail when no osds (#10257 Joao Eduardo Luis)
- mon: fix timeout rounds period (#10546 Joao Eduardo Luis)

- mon: ignore osd failures before up\_from (#10762 Dan van der Ster, Sage Weil)
- mon: paxos: reset accept timeout before writing to store (#10220 Joao Eduardo Luis)
- mon: return if fs exists on 'fs new' (Joao Eduardo Luis)
- mon: use EntityName when expanding profiles (#10844 Joao Eduardo Luis)
- mon: verify cross-service proposal preconditions (#10643 Joao Eduardo Luis)
- mon: wait for osdmon to be writeable when requesting proposal (#9794 Joao Eduardo Luis)
- mount.ceph: avoid spurious error message about /etc/mtab (#10351 Yan, Zheng)
- msg/simple: allow RESETSESSION when we forget an endpoint (#10080 Greg Farnum)
- msg/simple: discard delay queue before incoming queue (#9910 Sage Weil)
- osd: clear\_primary\_state when leaving Primary (#10059 Samuel Just)
- osd: do not ignore deleted pgs on startup (#10617 Sage Weil)
- osd: fix FileJournal wrap to get header out first (#10883 David Zafman)
- osd: fix PG leak in SnapTrimWQ (#10421 Kefu Chai)
- osd: fix journalq population in do\_read\_entry (#6003 Samuel Just)
- osd: fix operator== for op\_queue\_age\_hit and fs\_perf\_stat (#10259 Samuel Just)
- osd: fix rare assert after split (#10430 David Zafman)
- osd: get pgid ancestor from last\_map when building past intervals (#10430 David Zafman)
- osd: include rollback\_info\_trimmed\_to in {read,write}\_log (#10157 Samuel Just)
- osd: lock header\_lock in DBObjectMap::sync (#9891 Samuel Just)
- osd: requeue blocked op before flush it was blocked on (#10512 Sage Weil)
- osd: tolerate missing object between list and attr get on backfill (#10150 Samuel Just)
- osd: use correct atime for eviction decision (Xinze Chi)
- rgw: flush XML header on get ACL request (#10106 Yehuda Sadeh)
- rgw: index swift keys appropriately (#10471 Hemant Bruman, Yehuda Sadeh)
- rgw: send cancel for bucket index pending ops (#10770 Baijiaruo, Yehuda Sadeh)

- rgw: swift: support X\_Remove\_Container-Meta-{key} (#01475 Dmytro Iurchenko)

For more detailed information, see [the complete changelog](#).

## v0.87.1 Giant

---

This is the first (and possibly final) point release for Giant. Our focus on stability fixes will be directed towards Hammer and Firefly.

We recommend that all v0.87 Giant users upgrade to this release.

## Upgrading

---

- Due to a change in the Linux kernel version 3.18 and the limits of the FUSE interface, ceph-fuse needs be mounted as root on at least some systems. See issues #9997, #10277, and #10542 for details.

## Notable Changes

---

- build: disable stack-execute bit on assembler objects (#10114 Dan Mick)
- build: support boost 1.57.0 (#10688 Ken Dreyer)
- ceph-disk: fix dmcrypt file permissions (#9785 Loic Dachary)
- ceph-disk: run partprobe after zap, behave with partx or partprobe (#9665 #9721 Loic Dachary)
- cephfs-journal-tool: fix import for aged journals (#9977 John Spray)
- cephfs-journal-tool: fix journal import (#10025 John Spray)
- ceph-fuse: use remount to trim kernel dcache (#10277 Yan, Zheng)
- common: add cctid meta variable (#6228 Adam Crume)
- common: fix dump of shard for ghobject\_t (#10063 Loic Dachary)
- crush: fix bucket weight underflow (#9998 Paweł Sadowski)
- erasure-code: enforce chunk size alignment (#10211 Loic Dachary)
- erasure-code: regression test suite (#9420 Loic Dachary)
- erasure-code: relax caucy w restrictions (#10325 Loic Dachary)
- libcephfs,ceph-fuse: allow xattr caps on inject\_release\_failure (#9800 John Spray)

- libcephfs,ceph-fuse: fix cap flush tid comparison (#9869 Greg Farnum)
- libcephfs,ceph-fuse: new flag to indicated sorted dcache (#9178 Yan, Zheng)
- libcephfs,ceph-fuse: prune cache before reconnecting to MDS (Yan, Zheng)
- librados: limit number of in-flight read requests (#9854 Jason Dillaman)
- libradospy: fix thread shutdown (#8797 Dan Mick)
- libradosstriper: fix locking issue in truncate (#10129 Sebastien Ponce)
- librbd: complete pending ops before closing mage (#10299 Jason Dillaman)
- librbd: fix error path on image open failure (#10030 Jason Dillaman)
- librbd: gracefully handle deleted/renamed pools (#10270 Jason Dillaman)
- librbd: handle errors when creating ioctx while listing children (#10123 Jason Dillaman)
- mds: fix compat version in MClientSession (#9945 John Spray)
- mds: fix journaler write error handling (#10011 John Spray)
- mds: fix locking for file size recovery (#10229 Yan, Zheng)
- mds: handle heartbeat\_reset during shutdown (#10382 John Spray)
- mds: store backtrace for straydir (Yan, Zheng)
- mon: allow tiers for FS pools (#10135 John Spray)
- mon: fix caching of last\_epoch\_clean, osdmap trimming (#9987 Sage Weil)
- mon: fix 'fs ls' on peons (#10288 John Spray)
- mon: fix MDS health status from peons (#10151 John Spray)
- mon: fix paxos off-by-one (#9301 Sage Weil)
- msgr: simple: do not block on takeover while holding global lock (#9921 Greg Farnum)
- osd: deep scrub must not abort if hinfo is missing (#10018 Loic Dachary)
- osd: fix misdirected op detection (#9835 Sage Weil)
- osd: fix past\_interval display for acting (#9752 Loic Dachary)
- osd: fix PG peering backoff when behind on osdmmaps (#10431 Sage Weil)
- osd: handle no-op write with snapshot case (#10262 Sage Weil)

- osd: use fast-dispatch (Sage Weil, Greg Farnum)
- rados: fix write to /dev/null (Loic Dachary)
- radosgw-admin: create subuser when needed (#10103 Yehuda Sadeh)
- rbd: avoid invalidating aio\_write buffer during image import (#10590 Jason Dillaman)
- rbd: fix export with images > 2GB (Vicente Cheng)
- rgw: change multipart upload id magic (#10271 Georgios Dimitrakakis, Yehuda Sadeh)
- rgw: check keystone auth for S3 POST (#10062 Abhishek Lekshmanan)
- rgw: check timestamp for S3 keystone auth (#10062 Abhishek Lekshmanan)
- rgw: fix partial GET with swift (#10553 Yehuda Sadeh)
- rgw: fix quota disable (#9907 Dong Lei)
- rgw: fix rare corruption of object metadata on put (#9576 Yehuda Sadeh)
- rgw: fix S3 object copy content-type (#9478 Yehuda Sadeh)
- rgw: headers end with rn (#9254 Benedikt Fraunhofer, Yehuda Sadeh)
- rgw: remove swift user manifest DLO hash calculation (#9973 Yehuda Sadeh)
- rgw: return correct len when len is 0 (#9877 Yehuda Sadeh)
- rgw: return X-Timestamp field (#8911 Yehuda Sadeh)
- rgw: run radosgw as apache with systemd (#10125)
- rgw: sent ETag on S3 object copy (#9479 Yehuda Sadeh)
- rgw: sent HTTP status reason explicitly in fastcgi (Yehuda Sadeh)
- rgw: set length for keystone token validation (#7796 Mark Kirkwood, Yehuda Sadeh)
- rgw: set ulimit -n on sysvinit before starting daemon (#9587 Sage Weil)
- rgw: update bucket index on set\_attrs (#5595 Yehuda Sadeh)
- rgw: update swift subuser permission masks when authenticating (#9918 Yehuda Sadeh)
- rgw: URL decode HTTP query params correction (#10271 Georgios Dimitrakakis, Yehuda Sadeh)
- rgw: use cached attrs while reading object attrs (#10307 Yehuda Sadeh)

- rgw: use strict\_strtoll for content length (#10701 Axel Dunkel, Yehuda Sadeh)

For more detailed information, see [the complete changelog](#).

## v0.87 Giant

---

This release will form the basis for the stable release Giant, v0.87.x. Highlights for Giant include:

- *RADOS Performance*: a range of improvements have been made in the OSD and client-side librados code that improve the throughput on flash backends and improve parallelism and scaling on fast machines.
- *CephFS*: we have fixed a raft of bugs in CephFS and built some basic journal recovery and diagnostic tools. Stability and performance of single-MDS systems is vastly improved in Giant. Although we do not yet recommend CephFS for production deployments, we do encourage testing for non-critical workloads so that we can better gauge the feature, usability, performance, and stability gaps.
- *Local Recovery Codes*: the OSDs now support an erasure-coding scheme that stores some additional data blocks to reduce the IO required to recover from single OSD failures.
- *Degraded vs misplaced*: the Ceph health reports from ‘ceph -s’ and related commands now make a distinction between data that is degraded (there are fewer than the desired number of copies) and data that is misplaced (stored in the wrong location in the cluster). The distinction is important because the latter does not compromise data safety.
- *Tiering improvements*: we have made several improvements to the cache tiering implementation that improve performance. Most notably, objects are not promoted into the cache tier by a single read; they must be found to be sufficiently hot before that happens.
- *Monitor performance*: the monitors now perform writes to the local data store asynchronously, improving overall responsiveness.
- *Recovery tools*: the ceph\_objectstore\_tool is greatly expanded to allow manipulation of an individual OSDs data store for debugging and repair purposes. This is most heavily used by our QA infrastructure to exercise recovery code.

## Upgrade Sequencing

---

- If your existing cluster is running a version older than v0.80.x Firefly, please first upgrade to the latest Firefly release before moving on to Giant. We have not tested upgrades directly from Emperor, Dumpling, or older releases.

We have tested:

- Firefly to Giant
- Dumpling to Firefly to Giant

- Please upgrade daemons in the following order:

- i. Monitors
- ii. OSDs
- iii. MDSS and/or radosgw

Note that the relative ordering of OSDs and monitors should not matter, but we primarily tested upgrading monitors first.

## Upgrading from v0.80x Firefly

- The client-side caching for librbd is now enabled by default (`rbd cache = true`). A safety option (`rbd cache writethrough until flush = true`) is also enabled so that writeback caching is not used until the library observes a ‘flush’ command, indicating that the librbd users is passing that operation through from the guest VM. This avoids potential data loss when used with older versions of qemu that do not support flush.

```
leveldb_write_buffer_size = 8*1024*1024 = 33554432 // 8MB leveldb_cache_size = 512*1024*1204 = 536870912 //
512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

- The ‘rados getxattr ...’ command used to add a gratuitous newline to the attr value; it now does not.
- The `*_kb perf` counters on the monitor have been removed. These are replaced with a new set of `*_bytes` counters (e.g., `cluster_osd_kb` is replaced by `cluster_osd_bytes` ).
- The `rd_kb` and `wr_kb` fields in the JSON dumps for pool stats (accessed via the `ceph df detail -f json-pretty` and related commands) have been replaced with corresponding `*_bytes` fields. Similarly, the `total_space`, `total_used`, and `total_avail` fields are replaced with `total_bytes`, `total_used_bytes`, and `total_avail_bytes` fields.
- The `rados df --format=json` output `read_bytes` and `write_bytes` fields were incorrectly reporting ops; this is now fixed.

- The `rados df --format=json` output previously included `read_kb` and `write_kb` fields; these have been removed. Please use `read_bytes` and `write_bytes` instead (and divide by 1024 if appropriate).
- The experimental keyvaluestore-dev OSD backend had an on-disk format change that prevents existing OSD data from being upgraded. This affects developers and testers only.
- mon-specific and osd-specific leveldb options have been removed. From this point onward users should use the `leveldb_*` generic options and add the options in the appropriate sections of their configuration files. Monitors will still maintain the following monitor-specific defaults:

```
leveldb_write_buffer_size = 8*1024*1024 = 33554432 // 8MB leveldb_cache_size = 512*1024*1204 = 536870912 //
512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

- CephFS support for the legacy anchor table has finally been removed. Users with file systems created before firefly should ensure that inodes with multiple hard links are modified *prior* to the upgrade to ensure that the backtraces are written properly. For example:

```
1. sudo find /mnt/cephfs -type f -links +1 -exec touch \{\}\ \;
```

- We disallow nonsensical 'tier cache-mode' transitions. From this point onward, 'writeback' can only transition to 'forward' and 'forward' can transition to 1) 'writeback' if there are dirty objects, or 2) any if there are no dirty objects.

## Notable Changes since v0.86

- ceph-disk: use new udev rules for centos7/rhel7 (#9747 Loic Dachary)
- libcephfs-java: fix fstat mode (Noah Watkins)
- librados: fix deadlock when listing PG contents (Guang Yang)
- librados: misc fixes to the new threading model (#9582 #9706 #9845 #9873 Sage Weil)
- mds: fix inotable initialization (Henry C Chang)
- mds: gracefully handle unknown lock type in flock requests (Yan, Zheng)
- mon: add read-only, read-write, and role-definer roles (Joao Eduardo Luis)
- mon: fix mon cap checks (Joao Eduardo Luis)

- mon: misc fixes for new paxos async writes (#9635 Sage Weil)
- mon: set scrub timestamps on PG creation (#9496 Joao Eduardo Luis)
- osd: erasure code: fix buffer alignment (Janne Grunau, Loic Dachary)
- osd: fix alloc hint induced crashes on mixed clusters (#9419 David Zafman)
- osd: fix backfill reservation release on rejection (#9626, Samuel Just)
- osd: fix ioprio option parsing (#9676 #9677 Loic Dachary)
- osd: fix memory leak during snap trimming (#9113 Samuel Just)
- osd: misc peering and recovery fixes (#9614 #9696 #9731 #9718 #9821 #9875 Samuel Just, Guang Yang)

## Notable Changes since v0.80.x Firefly

---

- bash completion improvements (Wido den Hollander)
- brag: fixes, improvements (Loic Dachary)
- buffer: improve rebuild\_page\_aligned (Ma Jianpeng)
- build: fix build on alpha (Michael Cree, Dmitry Smirnov)
- build: fix CentOS 5 (Gerben Meijer)
- build: fix yasm check for x32 (Daniel Schepler, Sage Weil)
- ceph-brag: add tox tests (Alfredo Deza)
- ceph-conf: flush log on exit (Sage Weil)
- ceph.conf: update sample (Sebastien Han)
- ceph-dencoder: refactor build a bit to limit dependencies (Sage Weil, Dan Mick)
- ceph-disk: add Scientific Linux support (Dan van der Ster)
- ceph-disk: do not inadvertently create directories (Owne Synge)
- ceph-disk: fix dmcrypt support (Sage Weil)
- ceph-disk: fix dmcrypt support (Stephen Taylor)
- ceph-disk: handle corrupt volumes (Stuart Longlang)
- ceph-disk: linter cleanup, logging improvements (Alfredo Deza)
- ceph-disk: partprobe as needed (Eric Eastman)

- ceph-disk: show information about dmcrypt in ‘ceph-disk list’ output (Sage Weil)
- ceph-disk: use partition type UUIDs and blkid (Sage Weil)
- ceph: fix for non-default cluster names (#8944, Dan Mick)
- ceph-fuse, libcephfs: asok hooks for handling session resets, timeouts (Yan, Zheng)
- ceph-fuse, libcephfs: fix crash in trim\_caps (John Spray)
- ceph-fuse, libcephfs: improve cap trimming (John Spray)
- ceph-fuse, libcephfs: improve traceless reply handling (Sage Weil)
- ceph-fuse, libcephfs: virtual xattrs for rstat (Yan, Zheng)
- ceph\_objectstore\_tool: vastly improved and extended tool for working offline with OSD data stores (David Zafman)
- ceph.spec: many fixes (Erik Logtenberg, Boris Ranto, Dan Mick, Sandon Van Ness)
- ceph.spec: split out ceph-common package, other fixes (Sandon Van Ness)
- ceph\_test\_librbd\_fsx: fix RNG, make deterministic (Ilya Dryomov)
- cephtool: fix help (Yilong Zhao)
- cephtool: refactor and improve CLI tests (Joao Eduardo Luis)
- cephtool: test cleanup (Joao Eduardo Luis)
- clang build fixes (John Spray, Danny Al-Gaaf)
- client: improved MDS session dumps (John Spray)
- common: add config diff admin socket command (Joao Eduardo Luis)
- common: add rwlock assertion checks (Yehuda Sadeh)
- common: fix dup log messages (#9080, Sage Weil)
- common: perfcounters now use atomics and go faster (Sage Weil)
- config: support G, M, K, etc. suffixes (Joao Eduardo Luis)
- coverity cleanups (Danny Al-Gaaf)
- crush: clean up CrushWrapper interface (Xiaozi Chen)
- crush: include new tunables in dump (Sage Weil)
- crush: make ruleset ids unique (Xiaozi Chen, Loic Dachary)

- crush: only require rule features if the rule is used (#8963, Sage Weil)
- crushtool: send output to stdout, not stderr (Wido den Hollander)
- doc: cache tiering (John Wilkins)
- doc: CRUSH updates (John Wilkins)
- doc: document new upstream wireshark dissector (Kevin Cox)
- doc: improve manual install docs (Francois Lafont)
- doc: keystone integration docs (John Wilkins)
- doc: librados example fixes (Kevin Dalley)
- doc: many doc updates (John Wilkins)
- doc: many install doc updates (John Wilkins)
- doc: misc updates (John Wilkins, Loic Dachary, David Moreau Simard, Wido den Hollander, Volker Voigt, Alfredo Deza, Stephen Jahl, Dan van der Ster)
- doc: osd primary affinity (John Wilkins)
- doc: pool quotas (John Wilkins)
- doc: pre-flight doc improvements (Kevin Dalley)
- doc: switch to an unencumbered font (Ross Turk)
- doc: updated simple configuration guides (John Wilkins)
- doc: update erasure docs (Loic Dachary, Venky Shankar)
- doc: update openstack docs (Josh Durgin)
- filestore: disable use of XFS hint (buggy on old kernels) (Samuel Just)
- filestore: fix xattr spillout (Greg Farnum, Haomai Wang)
- fix hppa arch build (Dmitry Smirnov)
- fix i386 builds (Sage Weil)
- fix struct vs class inconsistencies (Thorsten Behrens)
- global: write pid file even when running in foreground (Alexandre Oliva)
- hadoop: improve tests (Huamin Chen, Greg Farnum, John Spray)
- hadoop: update hadoop tests for Hadoop 2.0 (Haumin Chen)
- init-ceph: continue starting other daemons on crush or mount failure (#8343, Sage Weil)

Weil)

- journaler: fix locking (Zheng, Yan)
- keyvaluestore: fix hint crash (#8381, Haomai Wang)
- keyvaluestore: header cache (Haomai Wang)
- libcephfs-java: build against older JNI headers (Greg Farnum)
- libcephfs-java: fix gcj-jdk build (Dmitry Smirnov)
- librados: fix crash on read op timeout (#9362 Matthias Kiefer, Sage Weil)
- librados: fix lock leaks in error paths (#9022, Paval Rallabhandi)
- librados: fix pool existence check (#8835, Pavan Rallabhandi)
- librados: fix rados\_pool\_list bounds checks (Sage Weil)
- librados: fix shutdown race (#9130 Sage Weil)
- librados: fix watch/notify test (#7934 David Zafman)
- librados: fix watch reregistration on acting set change (#9220 Samuel Just)
- librados: give Objecter fine-grained locks (Yehuda Sadeh, Sage Weil, John Spray)
- librados: lttng tracepoints (Adam Crume)
- librados, osd: return ETIMEDOUT on failed notify (Sage Weil)
- librados: pybind: fix reads when 0 is present (#9547 Mohammad Salehe)
- librados\_striper: striping library for librados (Sebastien Ponce)
- librbd, ceph-fuse: reduce cache flush overhead (Haomai Wang)
- librbd: check error code on cache invalidate (Josh Durgin)
- librbd: enable caching by default (Sage Weil)
- librbd: enforce cache size on read requests (Jason Dillaman)
- librbd: fix crash using clone of flattened image (#8845, Josh Durgin)
- librbd: fix error path when opening image (#8912, Josh Durgin)
- librbd: handle blacklisting during shutdown (#9105 John Spray)
- librbd: lttng tracepoints (Adam Crume)
- librbd: new libkrbd library for kernel map/unmap/showmapped (Ilya Dryomov)

- librbd: store and retrieve snapshot metadata based on id (Josh Durgin)
- libs3: update to latest (Danny Al-Gaaf)
- log: fix derr level (Joao Eduardo Luis)
- logrotate: fix osd log rotation on ubuntu (Sage Weil)
- lttng: tracing infrastructure (Noah Watkins, Adam Crume)
- mailmap: many updates (Loic Dachary)
- mailmap: updates (Loic Dachary, Abhishek Lekshmanan, M Ranga Swami Reddy)
- Makefile: fix out of source builds (Stefan Eilemann)
- many many coverity fixes, cleanups (Danny Al-Gaaf)
- mds: adapt to new Objecter locking, give types to all Contexts (John Spray)
- mds: add file system name, enabled flag (John Spray)
- mds: add internal health checks (John Spray)
- mds: add min/max UID for snapshot creation/deletion (#9029, Wido den Hollander)
- mds: avoid tight mon reconnect loop (#9428 Sage Weil)
- mds: boot refactor, cleanup (John Spray)
- mds: cephfs-journal-tool (John Spray)
- mds: fix crash killing sessions (#9173 John Spray)
- mds: fix ctime updates (#9514 Greg Farnum)
- mds: fix journal conversion with standby-replay (John Spray)
- mds: fix replay locking (Yan, Zheng)
- mds: fix standby-replay cache trimming (#8648 Zheng, Yan)
- mds: fix xattr bug triggered by ACLs (Yan, Zheng)
- mds: give perfcounters meaningful names (Sage Weil)
- mds: improve health reporting to monitor (John Spray)
- mds: improve Journaler on-disk format (John Spray)
- mds: improve journal locking (Zheng, Yan)
- mds, libcephfs: use client timestamp for mtime/ctime (Sage Weil)

- mds: make max file recoveries tunable (Sage Weil)
- mds: misc encoding improvements (John Spray)
- mds: misc fixes for multi-mds (Yan, Zheng)
- mds: multi-mds fixes (Yan, Zheng)
- mds: OPTracker integration, dump\_ops\_in\_flight (Greg Farnum)
- mds: prioritize file recovery when appropriate (Sage Weil)
- mds: refactor beacon, improve reliability (John Spray)
- mds: remove legacy anchor table (Yan, Zheng)
- mds: remove legacy discover ino (Yan, Zheng)
- mds: restart on EBLACKLISTED (John Spray)
- mds: separate inode recovery queue (John Spray)
- mds: session ls, evict commands (John Spray)
- mds: submit log events in async thread (Yan, Zheng)
- mds: track RECALL progress, report failure (#9284 John Spray)
- mds: update segment references during journal write (John Spray, Greg Farnum)
- mds: use client-provided timestamp for user-visible file metadata (Yan, Zheng)
- mds: use meaningful names for clients (John Spray)
- mds: validate journal header on load and save (John Spray)
- mds: warn clients which aren't revoking caps (Zheng, Yan, John Spray)
- misc build errors/warnings for Fedora 20 (Boris Ranto)
- misc build fixes for OS X (John Spray)
- misc cleanup (Christophe Courtaut)
- misc integer size cleanups (Kevin Cox)
- misc memory leaks, cleanups, fixes (Danny Al-Gaaf, Sahid Ferdjaoui)
- misc suse fixes (Danny Al-Gaaf)
- misc word size fixes (Kevin Cox)
- mon: add audit log for all admin commands (Joao Eduardo Luis)

- mon: add cluster fingerprint (Sage Weil)
- mon: add get-quota commands (Joao Eduardo Luis)
- mon: add ‘osd blocked-by’ command to easily see which OSDs are blocking peering progress (Sage Weil)
- mon: add ‘osd reweight-by-pg’ command (Sage Weil, Guang Yang)
- mon: add perfcounters for paxos operations (Sage Weil)
- mon: avoid creating unnecessary rule on pool create (#9304 Loic Dachary)
- monclient: fix hang (Sage Weil)
- mon: create default EC profile if needed (Loic Dachary)
- mon: do not create file system by default (John Spray)
- mon: do not spam log (Aanchal Agrawal, Sage Weil)
- mon: drop mon- and osd- specific leveldb options (Joao Eduardo Luis)
- mon: ec pool profile fixes (Loic Dachary)
- mon: fix bug when no auth keys are present (#8851, Joao Eduardo Luis)
- mon: fix ‘ceph df’ output for available space (Xiaoxi Chen)
- mon: fix compat version for MForward (Joao Eduardo Luis)
- mon: fix crash on loopback messages and paxos timeouts (#9062, Sage Weil)
- mon: fix default replication pool ruleset choice (#8373, John Spray)
- mon: fix divide by zero when pg\_num is adjusted before OSDs are added (#9101, Sage Weil)
- mon: fix double-free of old MOSDBoot (Sage Weil)
- mon: fix health down messages (Sage Weil)
- mon: fix occasional memory leak after session reset (#9176, Sage Weil)
- mon: fix op write latency perfcounter (#9217 Xinxin Shu)
- mon: fix ‘osd perf’ reported latency (#9269 Samuel Just)
- mon: fix quorum feature check (#8738, Greg Farnum)
- mon: fix ruleset/ruleid bugs (#9044, Loic Dachary)
- mon: fix set cache\_target\_full\_ratio (#8440, Geoffrey Hartz)

- mon: fix store check on startup (Joao Eduardo Luis)
- mon: include per-pool 'max avail' in df output (Sage Weil)
- mon: make paxos transaction commits asynchronous (Sage Weil)
- mon: make usage dumps in terms of bytes, not kB (Sage Weil)
- mon: 'osd crush reweight-subtree ...' (Sage Weil)
- mon, osd: relax client EC support requirements (Sage Weil)
- mon: preload erasure plugins (#9153 Loic Dachary)
- mon: prevent cache pools from being used directly by CephFS (#9435 John Spray)
- mon: prevent EC pools from being used with cephfs (Joao Eduardo Luis)
- mon: prevent implicit destruction of OSDs with 'osd setmaxosd ...' (#8865, Anand Bhat)
- mon: prevent nonsensical cache-mode transitions (Joao Eduardo Luis)
- mon: restore original weight when auto-marked out OSDs restart (Sage Weil)
- mon: restrict some pool properties to tiered pools (Joao Eduardo Luis)
- mon: some instrumentation (Sage Weil)
- mon: use msg header tid for MMonGetVersionReply (Ilya Dryomov)
- mon: use user-provided ruleset for replicated pool (Xiaoxi Chen)
- mon: verify all quorum members are contiguous at end of Paxos round (#9053, Sage Weil)
- mon: verify available disk space on startup (#9502 Joao Eduardo Luis)
- mon: verify erasure plugin version on load (Loic Dachary)
- msgr: avoid big lock when sending (most) messages (Greg Farnum)
- msgr: fix logged address (Yongyue Sun)
- msgr: misc locking fixes for fast dispatch (#8891, Sage Weil)
- msgr: refactor to cleanly separate SimpleMessenger implemenetation, move toward Connection-based calls (Matt Benjamin, Sage Wei)
- objecter: flag operations that are redirected by caching (Sage Weil)
- objectstore: clean up KeyValueDB interface for key/value backends (Sage Weil)
- osd: account for hit\_set\_archive bytes (Sage Weil)

- osd: add ability to prehash filestore directories (Guang Yang)
- osd: add ‘dump\_reservations’ admin socket command (Sage Weil)
- osd: add feature bit for erasure plugins (Loic Dachary)
- osd: add header cache for KeyValueStore (Haomai Wang)
- osd: add ISA erasure plugin table cache (Andreas-Joachim Peters)
- osd: add local\_mtime for use by cache agent (Zhiqiang Wang)
- osd: add local recovery code (LRC) erasure plugin (Loic Dachary)
- osd: add prototype KineticStore based on Seagate Kinetic (Josh Durgin)
- osd: add READFORWARD caching mode (Luis Pabon)
- osd: add superblock for KeyValueStore backend (Haomai Wang)
- osd: add support for Intel ISA-L erasure code library (Andreas-Joachim Peters)
- osd: allow map cache size to be adjusted at runtime (Sage Weil)
- osd: avoid refcounting overhead by passing a few things by ref (Somnath Roy)
- osd: avoid sharing PG info that is not durable (Samuel Just)
- osd: bound osdmap epoch skew between PGs (Sage Weil)
- osd: cache tier flushing fixes for snapped objects (Samuel Just)
- osd: cap hit\_set size (#9339 Samuel Just)
- osd: clean up shard\_id\_t, shard\_t (Loic Dachary)
- osd: clear FDCache on unlink (#8914 Loic Dachary)
- osd: clear slow request latency info on osd up/down (Sage Weil)
- osd: do not evict blocked objects (#9285 Zhiqiang Wang)
- osd: do not skip promote for write-ordered reads (#9064, Samuel Just)
- osd: fix agent early finish looping (David Zafman)
- osd: fix ambiguous encoding order for blacklisted clients (#9211, Sage Weil)
- osd: fix bogus assert during OSD shutdown (Sage Weil)
- osd: fix bug with long object names and rename (#8701, Sage Weil)
- osd: fix cache flush corner case for snapshotted objects (#9054, Samuel Just)

- osd: fix cache full -> not full requeueing (#8931, Sage Weil)
- osd: fix clone deletion case (#8334, Sam Just)
- osd: fix clone vs cache\_evict bug (#8629 Sage Weil)
- osd: fix connection reconnect race (Greg Farnum)
- osd: fix crash from duplicate backfill reservation (#8863 Sage Weil)
- osd: fix dead peer connection checks (#9295 Greg Farnum, Sage Weil)
- osd: fix discard of old/obsolete subop replies (#9259, Samuel Just)
- osd: fix discard of peer messages from previous intervals (Greg Farnum)
- osd: fix dump of open fds on EMFILE (Sage Weil)
- osd: fix dumps (Joao Eduardo Luis)
- osd: fix erasure-code lib initialization (Loic Dachary)
- osd: fix extent normalization (Adam Crume)
- osd: fix filestore removal corner case (#8332, Sam Just)
- osd: fix flush vs OpContext (Samuel Just)
- osd: fix gating of messages from old OSD instances (Greg Farnum)
- osd: fix hang waiting for osdmap (#8338, Greg Farnum)
- osd: fix interval check corner case during peering (#8104, Sam Just)
- osd: fix ISA erasure alignment (Loic Dachary, Andreas-Joachim Peters)
- osd: fix journal dump (Ma Jianpeng)
- osd: fix journal-less operation (Sage Weil)
- osd: fix keyvaluestore scrub (#8589 Haomai Wang)
- osd: fix keyvaluestore upgrade (Haomai Wang)
- osd: fix loopback msgr issue (Ma Jianpeng)
- osd: fix LSB release parsing (Danny Al-Gaaf)
- osd: fix MarkMeDown and other shutdown races (Sage Weil)
- osd: fix memstore bugs with collection\_move\_rename, lock ordering (Sage Weil)
- osd: fix min\_read\_recency\_for\_promote default on upgrade (Zhiqiang Wang)

- osd: fix mon feature bit requirements bug and resulting log spam (Sage Weil)
- osd: fix mount/remount sync race (#9144 Sage Weil)
- osd: fix PG object listing/ordering bug (Guang Yang)
- osd: fix PG stat errors with tiering (#9082, Sage Weil)
- osd: fix purged\_snap initialization on backfill (Sage Weil, Samuel Just, Dan van der Ster, Florian Haas)
- osd: fix race condition on object deletion (#9480 Somnath Roy)
- osd: fix recovery chunk size usage during EC recovery (Ma Jianpeng)
- osd: fix recovery reservation deadlock for EC pools (Samuel Just)
- osd: fix removal of old xattrs when overwriting chained xattrs (Ma Jianpeng)
- osd: fix requesting queueing on PG split (Samuel Just)
- osd: fix scrub vs cache bugs (Samuel Just)
- osd: fix snap object writeback from cache tier (#9054 Samuel Just)
- osd: fix trim of hitsets (Sage Weil)
- osd: force new xattrs into leveldb if fs returns E2BIG (#7779, Sage Weil)
- osd: implement alignment on chunk sizes (Loic Dachary)
- osd: improved backfill priorities (Sage Weil)
- osd: improve journal shutdown (Ma Jianpeng, Mark Kirkwood)
- osd: improve locking for KeyValueStore (Haomai Wang)
- osd: improve locking in OpTracker (Pavan Rallabhandi, Somnath Roy)
- osd: improve prioritization of recovery of degraded over misplaced objects (Sage Weil)
- osd: improve tiering agent arithmetic (Zhiqiang Wang, Sage Weil, Samuel Just)
- osd: include backend information in metadata reported to mon (Sage Weil)
- osd: locking, sharding, caching improvements in FileStore's FDCache (Somnath Roy, Greg Farnum)
- osd: lttng tracepoints for filestore (Noah Watkins)
- osd: make blacklist encoding deterministic (#9211 Sage Weil)
- osd: make tiering behave if hit\_sets aren't enabled (Sage Weil)

- osd: many important bug fixes (Samuel Just)
- osd: many many core fixes (Samuel Just)
- osd: many many important fixes (#8231 #8315 #9113 #9179 #9293 #9294 #9326 #9453 #9481 #9482 #9497 #9574 Samuel Just)
- osd: mark pools with incomplete clones (Sage Weil)
- osd: misc erasure code plugin fixes (Loic Dachary)
- osd: misc locking fixes for fast dispatch (Samuel Just, Ma Jianpeng)
- osd, mon: add rocksdb support (Xinxin Shu, Sage Weil)
- osd, mon: config sanity checks on start (Sage Weil, Joao Eduardo Luis)
- osd, mon: distinguish between “misplaced” and “degraded” objects in cluster health and PG state reporting (Sage Weil)
- osd, msgr: fast-dispatch of OSD ops (Greg Farnum, Samuel Just)
- osd, objecter: resend ops on last\_force\_op\_resend barrier; fix cache overlay op ordering (Sage Weil)
- osd: preload erasure plugins (#9153 Loic Dachary)
- osd: prevent old rados clients from using tiered pools (#8714, Sage Weil)
- osd: reduce OpTracker overhead (Somnath Roy)
- osd: refactor some ErasureCode functionality into command parent class (Loic Dachary)
- osd: remove obsolete classic scrub code (David Zafman)
- osd: scrub PGs with invalid stats (Sage Weil)
- osd: set configurable hard limits on object and xattr names (Sage Weil, Haomai Wang)
- osd: set rollback\_info\_completed on create (#8625, Samuel Just)
- osd: sharded threadpool to improve parallelism (Somnath Roy)
- osd: shard OpTracker to improve performance (Somnath Roy)
- osd: simple io prioritization for scrub (Sage Weil)
- osd: simple scrub throttling (Sage Weil)
- osd: simple snap trimmer throttle (Sage Weil)

- osd: tests for bench command (Loic Dachary)
- osd: trim old EC objects quickly; verify on scrub (Samuel Just)
- osd: use FIEMAP to inform copy\_range (Haomai Wang)
- osd: use local time for tiering decisions (Zhiqiang Wang)
- osd: use xfs hint less frequently (Ilya Dryomov)
- osd: verify erasure plugin version on load (Loic Dachary)
- osd: work around GCC 4.8 bug in journal code (Matt Benjamin)
- pybind/rados: fix small timeouts (John Spray)
- qa: xfstests updates (Ilya Dryomov)
- rados: allow setxattr value to be read from stdin (Sage Weil)
- rados bench: fix arg order (Kevin Dalley)
- rados: drop gratuitous n from getxattr command (Sage Weil)
- rados: fix bench write arithmetic (Jiangheng)
- rados: fix {read,write}\_ops values for df output (Sage Weil)
- rbd: add rbdfmap pre- and post post- hooks, fix misc bugs (Dmitry Smirnov)
- rbd-fuse: allow exposing single image (Stephen Taylor)
- rbd-fuse: fix unlink (Josh Durgin)
- rbd: improve option default behavior (Josh Durgin)
- rbd: parallelize rbd import, export (Jason Dillaman)
- rbd: rbd-replay utility to replay captured rbd workload traces (Adam Crume)
- rbd: use write-back (not write-through) when caching is enabled (Jason Dillaman)
- removed mkcephfs (deprecated since dumpling)
- rest-api: fix help (Ailing Zhang)
- rgw: add civetweb as default frontend on port 7490 (#9013 Yehuda Sadeh)
- rgw: add -min-rewrite-stripe-size for object restriper (Yehuda Sadeh)
- rgw: add powerdns hook for dynamic DNS for global clusters (Wido den Hollander)
- rgw: add S3 bucket get location operation (Abhishek Lekshmanan)

- rgw: allow : in S3 access key (Roman Haritonov)
- rgw: automatically align writes to EC pool (#8442, Yehuda Sadeh)
- rgw: bucket link uses instance id (Yehuda Sadeh)
- rgw: cache bucket info (Yehuda Sadeh)
- rgw: cache decoded user info (Yehuda Sadeh)
- rgw: check entity permission for put\_metadata (#8428, Yehuda Sadeh)
- rgw: copy object data is target bucket is in a different pool (#9039, Yehuda Sadeh)
- rgw: do not try to authenticate CORS preflight requests (#8718, Robert Hubbard, Yehuda Sadeh)
- rgw: fix admin create user op (#8583 Ray Lv)
- rgw: fix civetweb URL decoding (#8621, Yehuda Sadeh)
- rgw: fix crash on swift CORS preflight request (#8586, Yehuda Sadeh)
- rgw: fix log filename suffix (#9353 Alexandre Marangone)
- rgw: fix memory leak following chunk read error (Yehuda Sadeh)
- rgw: fix memory leaks (Andrey Kuznetsov)
- rgw: fix multipart object attr regression (#8452, Yehuda Sadeh)
- rgw: fix multipart upload (#8846, Silvain Munaut, Yehuda Sadeh)
- rgw: fix radosgw-admin 'show log' command (#8553, Yehuda Sadeh)
- rgw: fix removal of objects during object creation (Patrycja Szablowska, Yehuda Sadeh)
- rgw: fix striping for copied objects (#9089, Yehuda Sadeh)
- rgw: fix test for identify whether an object has a tail (#9226, Yehuda Sadeh)
- rgw: fix URL decoding (#8702, Brian Rak)
- rgw: fix URL escaping (Yehuda Sadeh)
- rgw: fix usage (Abhishek Lekshmanan)
- rgw: fix user manifest (Yehuda Sadeh)
- rgw: fix when stripe size is not a multiple of chunk size (#8937, Yehuda Sadeh)
- rgw: handle empty extra pool name (Yehuda Sadeh)

- rgw: improve civetweb logging (Yehuda Sadeh)
- rgw: improve delimited listing of bucket, misc fixes (Yehuda Sadeh)
- rgw: improve -h (Abhishek Lekshmanan)
- rgw: many fixes for civetweb (Yehuda Sadeh)
- rgw: misc civetweb fixes (Yehuda Sadeh)
- rgw: misc civetweb frontend fixes (Yehuda Sadeh)
- rgw: object and bucket rewrite functions to allow restriping old objects (Yehuda Sadeh)
- rgw: powerdns backend for global namespaces (Wido den Hollander)
- rgw: prevent multiobject PUT race (Yehuda Sadeh)
- rgw: send user manifest header (Yehuda Sadeh)
- rgw: subuser creation fixes (#8587 Yehuda Sadeh)
- rgw: use systemd-run from sysvinit script (JuanJose Galvez)
- rpm: do not restart daemons on upgrade (Alfredo Deza)
- rpm: misc packaging fixes for rhel7 (Sandon Van Ness)
- rpm: split ceph-common from ceph (Sandon Van Ness)
- systemd: initial systemd config files (Federico Simoncelli)
- systemd: wrap started daemons in new systemd environment (Sage Weil, Dan Mick)
- sysvinit: add support for non-default cluster names (Alfredo Deza)
- sysvinit: less sensitive to failures (Sage Weil)
- test\_librbd\_fsx: test krbd as well as librbd (Ilya Dryomov)
- unit test improvements (Loic Dachary)
- upstart: increase max open files limit (Sage Weil)
- vstart.sh: fix/improve rgw support (Luis Pabon, Abhishek Lekshmanan)

## v0.86

This is a release candidate for Giant, which will hopefully be out in another week or two. We did a feature freeze about a month ago and since then have been doing only stabilization and bug fixing (and a handful of low-risk enhancements). A fair bit of

new functionality went into the final sprint, but it's baked for quite a while now and we're feeling pretty good about it.

Major items include:

- librados locking refactor to improve scaling and client performance
- local recovery code (LRC) erasure code plugin to trade some additional storage overhead for improved recovery performance
- LTTNG tracing framework, with initial tracepoints in librados, librbd, and the OSD FileStore backend
- separate monitor audit log for all administrative commands
- asynchronous monitor transaction commits to reduce the impact on monitor read requests while processing updates
- low-level tool for working with individual OSD data stores for debugging, recovery, and testing
- many MDS improvements (bug fixes, health reporting)

There are still a handful of known bugs in this release, but nothing severe enough to prevent a release. By and large we are pretty pleased with the stability and expect the final Giant release to be quite reliable.

Please try this out on your non-production clusters for a preview

## Notable Changes

---

- buffer: improve rebuild\_page\_aligned (Ma Jianpeng)
- build: fix CentOS 5 (Gerben Meijer)
- build: fix build on alpha (Michael Cree, Dmitry Smirnov)
- build: fix yasm check for x32 (Daniel Schepler, Sage Weil)
- ceph-disk: add Scientific Linux support (Dan van der Ster)
- ceph-fuse, libcephfs: fix crash in trim\_caps (John Spray)
- ceph-fuse, libcephfs: improve cap trimming (John Spray)
- ceph-fuse, libcephfs: virtual xattrs for rstat (Yan, Zheng)
- ceph.conf: update sample (Sebastien Han)
- ceph.spec: many fixes (Erik Logtenberg, Boris Ranto, Dan Mick, Sandon Van Ness)
- ceph\_objectstore\_tool: vastly improved and extended tool for working offline with

## OSD data stores (David Zafman)

- common: add config diff admin socket command (Joao Eduardo Luis)
- common: add rwlock assertion checks (Yehuda Sadeh)
- crush: clean up Crushwrapper interface (Xiaoaxi Chen)
- crush: make ruleset ids unique (Xiaoxi Chen, Loic Dachary)
- doc: improve manual install docs (Francois Lafont)
- doc: misc updates (John Wilkins, Loic Dachary, David Moreau Simard, Wido den Hollander, Volker Voigt, Alfredo Deza, Stephen Jahl, Dan van der Ster)
- global: write pid file even when running in foreground (Alexandre Oliva)
- hadoop: improve tests (Huamin Chen, Greg Farnum, John Spray)
- journaler: fix locking (Zheng, Yan)
- librados, osd: return ETIMEDOUT on failed notify (Sage Weil)
- librados: fix crash on read op timeout (#9362 Matthias Kiefer, Sage Weil)
- librados: fix shutdown race (#9130 Sage Weil)
- librados: fix watch reregistration on acting set change (#9220 Samuel Just)
- librados: fix watch/notify test (#7934 David Zafman)
- librados: give Objecter fine-grained locks (Yehuda Sadeh, Sage Weil, John Spray)
- librados: lttng tracepoints (Adam Crume)
- librados: pybind: fix reads when 0 is present (#9547 Mohammad Salehe)
- librbd: enforce cache size on read requests (Jason Dillaman)
- librbd: handle blacklisting during shutdown (#9105 John Spray)
- librbd: lttng tracepoints (Adam Crume)
- lttng: tracing infrastructure (Noah Watkins, Adam Crume)
- mailmap: updates (Loic Dachary, Abhishek Lekshmanan, M Ranga Swami Reddy)
- many many coverity fixes, cleanups (Danny Al-Gaaf)
- mds: adapt to new Objecter locking, give types to all Contexts (John Spray)
- mds: add internal health checks (John Spray)
- mds: avoid tight mon reconnect loop (#9428 Sage Weil)

- mds: fix crash killing sessions (#9173 John Spray)
- mds: fix ctime updates (#9514 Greg Farnum)
- mds: fix replay locking (Yan, Zheng)
- mds: fix standby-replay cache trimming (#8648 Zheng, Yan)
- mds: give perfcounters meaningful names (Sage Weil)
- mds: improve health reporting to monitor (John Spray)
- mds: improve journal locking (Zheng, Yan)
- mds: make max file recoveries tunable (Sage Weil)
- mds: prioritize file recovery when appropriate (Sage Weil)
- mds: refactor beacon, improve reliability (John Spray)
- mds: restart on EBLACKLISTED (John Spray)
- mds: track RECALL progress, report failure (#9284 John Spray)
- mds: update segment references during journal write (John Spray, Greg Farnum)
- mds: use meaningful names for clients (John Spray)
- mds: warn clients which aren't revoking caps (Zheng, Yan, John Spray)
- mon: add 'osd reweight-by-pg' command (Sage Weil, Guang Yang)
- mon: add audit log for all admin commands (Joao Eduardo Luis)
- mon: add cluster fingerprint (Sage Weil)
- mon: avoid creating unnecessary rule on pool create (#9304 Loic Dachary)
- mon: do not spam log (Aanchal Agrawal, Sage Weil)
- mon: fix 'osd perf' reported latency (#9269 Samuel Just)
- mon: fix double-free of old MOSDBoot (Sage Weil)
- mon: fix op write latency perfcounter (#9217 Xinxin Shu)
- mon: fix store check on startup (Joao Eduardo Luis)
- mon: make paxos transaction commits asynchronous (Sage Weil)
- mon: preload erasure plugins (#9153 Loic Dachary)
- mon: prevent cache pools from being used directly by CephFS (#9435 John Spray)

- mon: use user-provided ruleset for replicated pool (Xiaoxi Chen)
- mon: verify available disk space on startup (#9502 Joao Eduardo Luis)
- mon: verify erasure plugin version on load (Loic Dachary)
- msgr: fix logged address (Yongyue Sun)
- osd: account for hit\_set\_archive bytes (Sage Weil)
- osd: add ISA erasure plugin table cache (Andreas-Joachim Peters)
- osd: add ability to prehash filestore directories (Guang Yang)
- osd: add feature bit for erasure plugins (Loic Dachary)
- osd: add local recovery code (LRC) erasure plugin (Loic Dachary)
- osd: cap hit\_set size (#9339 Samuel Just)
- osd: clear FDCache on unlink (#8914 Loic Dachary)
- osd: do not evict blocked objects (#9285 Zhiqiang Wang)
- osd: fix ISA erasure alignment (Loic Dachary, Andreas-Joachim Peters)
- osd: fix clone vs cache\_evict bug (#8629 Sage Weil)
- osd: fix crash from duplicate backfill reservation (#8863 Sage Weil)
- osd: fix dead peer connection checks (#9295 Greg Farnum, Sage Weil)
- osd: fix keyvaluestore scrub (#8589 Haomai Wang)
- osd: fix keyvaluestore upgrade (Haomai Wang)
- osd: fix min\_read\_recency\_for\_promote default on upgrade (Zhiqiang Wang)
- osd: fix mount/remount sync race (#9144 Sage Weil)
- osd: fix purged\_snap initialization on backfill (Sage Weil, Samuel Just, Dan van der Ster, Florian Haas)
- osd: fix race condition on object deletion (#9480 Somnath Roy)
- osd: fix snap object writeback from cache tier (#9054 Samuel Just)
- osd: improve journal shutdown (Ma Jianpeng, Mark Kirkwood)
- osd: improve locking in OpTracker (Pavan Rallabhandi, Somnath Roy)
- osd: improve tiering agent arithmetic (Zhiqiang Wang, Sage Weil, Samuel Just)
- osd: lttng tracepoints for filestore (Noah Watkins)

- osd: make blacklist encoding deterministic (#9211 Sage Weil)
- osd: many many important fixes (#8231 #8315 #9113 #9179 #9293 #9294 #9326 #9453 #9481 #9482 #9497 #9574 Samuel Just)
- osd: misc erasure code plugin fixes (Loic Dachary)
- osd: preload erasure plugins (#9153 Loic Dachary)
- osd: shard OpTracker to improve performance (Somnath Roy)
- osd: use local time for tiering decisions (Zhiqiang Wang)
- osd: verify erasure plugin version on load (Loic Dachary)
- rados: fix bench write arithmetic (Jiangheng)
- rbd: parallelize rbd import, export (Jason Dillaman)
- rbd: rbd-replay utility to replay captured rbd workload traces (Adam Crume)
- rbd: use write-back (not write-through) when caching is enabled (Jason Dillaman)
- rgw: add S3 bucket get location operation (Abhishek Lekshmanan)
- rgw: add civetweb as default frontend on port 7490 (#9013 Yehuda Sadeh)
- rgw: allow : in S3 access key (Roman Haritonov)
- rgw: fix admin create user op (#8583 Ray Lv)
- rgw: fix log filename suffix (#9353 Alexandre Marangone)
- rgw: fix usage (Abhishek Lekshmanan)
- rgw: many fixes for civetweb (Yehuda Sadeh)
- rgw: subuser creation fixes (#8587 Yehuda Sadeh)
- rgw: use systemd-run from sysvinit script (JuanJose Galvez)
- unit test improvements (Loic Dachary)
- vstart.sh: fix/improve rgw support (Luis Pabon, Abhishek Lekshmanan)

## v0.85

---

This is the second-to-last development release before Giant that contains new functionality. The big items to land during this cycle are the messenger refactoring from Matt Benjmain that lays some groundwork for RDMA support, a performance improvement series from SanDisk that improves performance on SSDs, lots of improvements to our new standalone civetweb-based RGW frontend, and a new ‘osd

blocked-by' mon command that allows admins to easily identify which OSDs are blocking peering progress. The other big change is that the OSDs and Monitors now distinguish between "misplaced" and "degraded" objects: the latter means there are fewer copies than we'd like, while the former simply means they are not stored in the locations where we want them to be.

Also of note is a change to librbd that enables client-side caching by default. This is coupled with another option that makes the cache write-through until a "flush" operation is observed: this implies that the librbd user (usually a VM guest OS) supports barriers and flush and that it is safe for the cache to switch into writeback mode without compromising data safety or integrity. It has long been recommended practice that these options be enabled (e.g., in OpenStack environments) but until now it has not been the default.

We have frozen the tree for the looming Giant release, and the next development release will be a release candidate with a final batch of new functionality.

## Upgrading

---

- The client-side caching for librbd is now enabled by default (rbd cache = true). A safety option (rbd cache writethrough until flush = true) is also enabled so that writeback caching is not used until the library observes a 'flush' command, indicating that the librbd user is passing that operation through from the guest VM. This avoids potential data loss when used with older versions of qemu that do not support flush.

```
leveldb_write_buffer_size = 32*1024*1024 = 33554432 // 32MB leveldb_cache_size = 512*1024*1204 = 536870912
// 512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

- The 'rados getxattr ...' command used to add a gratuitous newline to the attr value; it now does not.

## Notable Changes

---

- ceph-disk: do not inadvertently create directories (Owne Synge)
- ceph-disk: fix dmcrypt support (Sage Weil)
- ceph-disk: linter cleanup, logging improvements (Alfredo Deza)
- ceph-disk: show information about dmcrypt in 'ceph-disk list' output (Sage Weil)
- ceph-disk: use partition type UUIDs and blkid (Sage Weil)

- ceph: fix for non-default cluster names (#8944, Dan Mick)
- doc: document new upstream wireshark dissector (Kevin Cox)
- doc: many install doc updates (John Wilkins)
- librados: fix lock leaks in error paths (#9022, Paval Rallabhandi)
- librados: fix pool existence check (#8835, Pavan Rallabhandi)
- librbd: enable caching by default (Sage Weil)
- librbd: fix crash using clone of flattened image (#8845, Josh Durgin)
- librbd: store and retrieve snapshot metadata based on id (Josh Durgin)
- mailmap: many updates (Loic Dachary)
- mds: add min/max UID for snapshot creation/deletion (#9029, Wido den Hollander)
- misc build errors/warnings for Fedora 20 (Boris Ranto)
- mon: add ‘osd blocked-by’ command to easily see which OSDs are blocking peering progress (Sage Weil)
- mon: add perfcounters for paxos operations (Sage Weil)
- mon: create default EC profile if needed (Loic Dachary)
- mon: fix crash on loopback messages and paxos timeouts (#9062, Sage Weil)
- mon: fix divide by zero when pg\_num is adjusted before OSDs are added (#9101, Sage Weil)
- mon: fix occasional memory leak after session reset (#9176, Sage Weil)
- mon: fix ruleset/ruleid bugs (#9044, Loic Dachary)
- mon: make usage dumps in terms of bytes, not kB (Sage Weil)
- mon: prevent implicit destruction of OSDs with ‘osd setmaxosd ...’ (#8865, Anand Bhat)
- mon: verify all quorum members are contiguous at end of Paxos round (#9053, Sage Weil)
- msgr: refactor to cleanly separate SimpleMessenger implemenetation, move toward Connection-based calls (Matt Benjamin, Sage Wei)
- objectstore: clean up KeyValueDB interface for key/value backends (Sage Weil)
- osd: add local\_mtime for use by cache agent (Zhiqiang Wang)

- osd: add superblock for KeyValueStore backend (Haomai Wang)
- osd: add support for Intel ISA-L erasure code library (Andreas-Joachim Peters)
- osd: do not skip promote for write-ordered reads (#9064, Samuel Just)
- osd: fix ambiguous encoding order for blacklisted clients (#9211, Sage Weil)
- osd: fix cache flush corner case for snapshotted objects (#9054, Samuel Just)
- osd: fix discard of old/obsolete subop replies (#9259, Samuel Just)
- osd: fix discard of peer messages from previous intervals (Greg Farnum)
- osd: fix dump of open fds on EMFILE (Sage Weil)
- osd: fix journal dump (Ma Jianpeng)
- osd: fix mon feature bit requirements bug and resulting log spam (Sage Weil)
- osd: fix recovery chunk size usage during EC recovery (Ma Jianpeng)
- osd: fix recovery reservation deadlock for EC pools (Samuel Just)
- osd: fix removal of old xattrs when overwriting chained xattrs (Ma Jianpeng)
- osd: fix requesting queueing on PG split (Samuel Just)
- osd: force new xattrs into leveldb if fs returns E2BIG (#7779, Sage Weil)
- osd: implement alignment on chunk sizes (Loic Dachary)
- osd: improve prioritization of recovery of degraded over misplaced objects (Sage Weil)
- osd: locking, sharding, caching improvements in FileStore's FDCache (Somnath Roy, Greg Farnum)
- osd: many important bug fixes (Samuel Just)
- osd, mon: add rocksdb support (Xinxin Shu, Sage Weil)
- osd, mon: distinguish between "misplaced" and "degraded" objects in cluster health and PG state reporting (Sage Weil)
- osd: refactor some ErasureCode functionality into command parent class (Loic Dachary)
- osd: set rollback\_info\_completed on create (#8625, Samuel Just)
- rados: allow setxattr value to be read from stdin (Sage Weil)
- rados: drop gratuitous n from getxattr command (Sage Weil)

- rgw: add `-min-rewrite-stripe-size` for object restriper (Yehuda Sadeh)
- rgw: add powerdns hook for dynamic DNS for global clusters (Wido den Hollander)
- rgw: copy object data is target bucket is in a different pool (#9039, Yehuda Sadeh)
- rgw: do not try to authenticate CORS preflight requests (#8718, Robert Hubbard, Yehuda Sadeh)
- rgw: fix civetweb URL decoding (#8621, Yehuda Sadeh)
- rgw: fix removal of objects during object creation (Patrycja Szablowska, Yehuda Sadeh)
- rgw: fix striping for copied objects (#9089, Yehuda Sadeh)
- rgw: fix test for identify whether an object has a tail (#9226, Yehuda Sadeh)
- rgw: fix when stripe size is not a multiple of chunk size (#8937, Yehuda Sadeh)
- rgw: improve civetweb logging (Yehuda Sadeh)
- rgw: misc civetweb frontend fixes (Yehuda Sadeh)
- sysvinit: add support for non-default cluster names (Alfredo Deza)

## v0.84

---

The next Ceph development release is here! This release contains several meaty items, including some MDS improvements for journaling, the ability to remove the CephFS file system (and name it), several mon cleanups with tiered pools, several OSD performance branches, a new “read forward” RADOS caching mode, a prototype Kinetic OSD backend, and various radosgw improvements (especially with the new standalone civetweb frontend). And there are a zillion OSD bug fixes. Things are looking pretty good for the Giant release that is coming up in the next month.

## Upgrading

---

- The `*_kb` counters on the monitor have been removed. These are replaced with a new set of `*_bytes` counters (e.g., `cluster_osd_kb` is replaced by `cluster_osd_bytes`).
- The `rd_kb` and `wr_kb` fields in the JSON dumps for pool stats (accessed via the `ceph df detail -f json-pretty` and related commands) have been replaced with corresponding `*_bytes` fields. Similarly, the `total_space`, `total_used`, and `total_avail` fields are replaced with `total_bytes`, `total_used_bytes`, and `total_avail_bytes` fields.

- The `rados df --format=json` output `read_bytes` and `write_bytes` fields were incorrectly reporting ops; this is now fixed.
- The `rados df --format=json` output previously included `read_kb` and `write_kb` fields; these have been removed. Please use `read_bytes` and `write_bytes` instead (and divide by 1024 if appropriate).

## Notable Changes

---

- ceph-conf: flush log on exit (Sage Weil)
- ceph-dencoder: refactor build a bit to limit dependencies (Sage Weil, Dan Mick)
- ceph.spec: split out ceph-common package, other fixes (Sandon Van Ness)
- ceph\_test\_librbd\_fsx: fix RNG, make deterministic (Ilya Dryomov)
- cephtool: refactor and improve CLI tests (Joao Eduardo Luis)
- client: improved MDS session dumps (John Spray)
- common: fix dup log messages (#9080, Sage Weil)
- crush: include new tunables in dump (Sage Weil)
- crush: only require rule features if the rule is used (#8963, Sage Weil)
- crushtool: send output to stdout, not stderr (Wido den Hollander)
- fix i386 builds (Sage Weil)
- fix struct vs class inconsistencies (Thorsten Behrens)
- hadoop: update hadoop tests for Hadoop 2.0 (Haumin Chen)
- librbd, ceph-fuse: reduce cache flush overhead (Haomai Wang)
- librbd: fix error path when opening image (#8912, Josh Durgin)
- mds: add file system name, enabled flag (John Spray)
- mds: boot refactor, cleanup (John Spray)
- mds: fix journal conversion with standby-replay (John Spray)
- mds: separate inode recovery queue (John Spray)
- mds: session ls, evict commands (John Spray)
- mds: submit log events in async thread (Yan, Zheng)
- mds: use client-provided timestamp for user-visible file metadata (Yan, Zheng)

- mds: validate journal header on load and save (John Spray)
- misc build fixes for OS X (John Spray)
- misc integer size cleanups (Kevin Cox)
- mon: add get-quota commands (Joao Eduardo Luis)
- mon: do not create file system by default (John Spray)
- mon: fix 'ceph df' output for available space (Xiaoxi Chen)
- mon: fix bug when no auth keys are present (#8851, Joao Eduardo Luis)
- mon: fix compat version for MForward (Joao Eduardo Luis)
- mon: restrict some pool properties to tiered pools (Joao Eduardo Luis)
- msgr: misc locking fixes for fast dispatch (#8891, Sage Weil)
- osd: add 'dump\_reservations' admin socket command (Sage Weil)
- osd: add READFORWARD caching mode (Luis Pabon)
- osd: add header cache for KeyValueStore (Haomai Wang)
- osd: add prototype KineticStore based on Seagate Kinetic (Josh Durgin)
- osd: allow map cache size to be adjusted at runtime (Sage Weil)
- osd: avoid refcounting overhead by passing a few things by ref (Somnath Roy)
- osd: avoid sharing PG info that is not durable (Samuel Just)
- osd: clear slow request latency info on osd up/down (Sage Weil)
- osd: fix PG object listing/ordering bug (Guang Yang)
- osd: fix PG stat errors with tiering (#9082, Sage Weil)
- osd: fix bug with long object names and rename (#8701, Sage Weil)
- osd: fix cache full -> not full requeueing (#8931, Sage Weil)
- osd: fix gating of messages from old OSD instances (Greg Farnum)
- osd: fix memstore bugs with collection\_move\_rename, lock ordering (Sage Weil)
- osd: improve locking for KeyValueStore (Haomai Wang)
- osd: make tiering behave if hit\_sets aren't enabled (Sage Weil)
- osd: mark pools with incomplete clones (Sage Weil)

- osd: misc locking fixes for fast dispatch (Samuel Just, Ma Jianpeng)
- osd: prevent old rados clients from using tiered pools (#8714, Sage Weil)
- osd: reduce OpTracker overhead (Somnath Roy)
- osd: set configurable hard limits on object and xattr names (Sage Weil, Haomai Wang)
- osd: trim old EC objects quickly; verify on scrub (Samuel Just)
- osd: work around GCC 4.8 bug in journal code (Matt Benjamin)
- rados bench: fix arg order (Kevin Dalley)
- rados: fix {read,write}\_ops values for df output (Sage Weil)
- rbd: add rbimap pre- and post post- hooks, fix misc bugs (Dmitry Smirnov)
- rbd: improve option default behavior (Josh Durgin)
- rgw: automatically align writes to EC pool (#8442, Yehuda Sadeh)
- rgw: fix crash on swift CORS preflight request (#8586, Yehuda Sadeh)
- rgw: fix memory leaks (Andrey Kuznetsov)
- rgw: fix multipart upload (#8846, Silvain Munaut, Yehuda Sadeh)
- rgw: improve -h (Abhishek Lekshmanan)
- rgw: improve delimited listing of bucket, misc fixes (Yehuda Sadeh)
- rgw: misc civetweb fixes (Yehuda Sadeh)
- rgw: powerdns backend for global namespaces (Wido den Hollander)
- systemd: initial systemd config files (Federico Simoncelli)

## v0.83

---

Another Ceph development release! This has been a longer cycle, so there has been quite a bit of bug fixing and stabilization in this round. There is also a bunch of packaging fixes for RPM distros (RHEL/CentOS, Fedora, and SUSE) and for systemd. We've also added a new librados-striper library from Sébastien Ponce that provides a generic striping API for applications to code to.

## Upgrading

---

- The experimental keyvaluestore-dev OSD backend had an on-disk format change that

prevents existing OSD data from being upgraded. This affects developers and testers only.

- mon-specific and osd-specific leveldb options have been removed. From this point onward users should use the leveldb\_\* generic options and add the options in the appropriate sections of their configuration files. Monitors will still maintain the following monitor-specific defaults:

```
leveldb_write_buffer_size = 32*1024*1024 = 33554432 // 32MB leveldb_cache_size = 512*1024*1204 = 536870912
// 512MB leveldb_block_size = 64*1024 = 65536 // 64KB leveldb_compression = false leveldb_log = ""
```

OSDs will still maintain the following osd-specific defaults:

```
leveldb_log = ""
```

## Notable Changes

---

- ceph-disk: fix dmcrypt support (Stephen Taylor)
- cephtool: fix help (Yilong Zhao)
- cephtool: test cleanup (Joao Eduardo Luis)
- doc: librados example fixes (Kevin Dalley)
- doc: many doc updates (John Wilkins)
- doc: update erasure docs (Loic Dachary, Venky Shankar)
- filestore: disable use of XFS hint (buggy on old kernels) (Samuel Just)
- filestore: fix xattr spillout (Greg Farnum, Haomai Wang)
- keyvaluestore: header cache (Haomai Wang)
- librados\_striper: striping library for librados (Sebastien Ponce)
- libs3: update to latest (Danny Al-Gaaf)
- log: fix derr level (Joao Eduardo Luis)
- logrotate: fix osd log rotation on ubuntu (Sage Weil)
- mds: fix xattr bug triggered by ACLs (Yan, Zheng)
- misc memory leaks, cleanups, fixes (Danny Al-Gaaf, Sahid Ferdjaoui)
- misc suse fixes (Danny Al-Gaaf)
- misc word size fixes (Kevin Cox)

- mon: drop mon- and osd- specific leveldb options (Joao Eduardo Luis)
- mon: ec pool profile fixes (Loic Dachary)
- mon: fix health down messages (Sage Weil)
- mon: fix quorum feature check (#8738, Greg Farnum)
- mon: ‘osd crush reweight-subtree ...’ (Sage Weil)
- mon, osd: relax client EC support requirements (Sage Weil)
- mon: some instrumentation (Sage Weil)
- objecter: flag operations that are redirected by caching (Sage Weil)
- osd: clean up shard\_id\_t, shard\_t (Loic Dachary)
- osd: fix connection reconnect race (Greg Farnum)
- osd: fix dumps (Joao Eduardo Luis)
- osd: fix erasure-code lib initialization (Loic Dachary)
- osd: fix extent normalization (Adam Crume)
- osd: fix loopback msgr issue (Ma Jianpeng)
- osd: fix LSB release parsing (Danny Al-Gaaf)
- osd: improved backfill priorities (Sage Weil)
- osd: many many core fixes (Samuel Just)
- osd, mon: config sanity checks on start (Sage Weil, Joao Eduardo Luis)
- osd: sharded threadpool to improve parallelism (Somnath Roy)
- osd: simple io prioritization for scrub (Sage Weil)
- osd: simple scrub throttling (Sage Weil)
- osd: tests for bench command (Loic Dachary)
- osd: use xfs hint less frequently (Ilya Dryomov)
- pybind/rados: fix small timeouts (John Spray)
- qa: xfstests updates (Ilya Dryomov)
- rgw: cache bucket info (Yehuda Sadeh)
- rgw: cache decoded user info (Yehuda Sadeh)

- rgw: fix multipart object attr regression (#8452, Yehuda Sadeh)
- rgw: fix radosgw-admin ‘show log’ command (#8553, Yehuda Sadeh)
- rgw: fix URL decoding (#8702, Brian Rak)
- rgw: handle empty extra pool name (Yehuda Sadeh)
- rpm: do not restart daemons on upgrade (Alfredo Deza)
- rpm: misc packaging fixes for rhel7 (Sandon Van Ness)
- rpm: split ceph-common from ceph (Sandon Van Ness)
- systemd: wrap started daemons in new systemd environment (Sage Weil, Dan Mick)
- sysvinit: less sensitive to failures (Sage Weil)
- upstart: increase max open files limit (Sage Weil)

## v0.82

---

This is the second post-firefly development release. It includes a range of bug fixes and some usability improvements. There are some MDS debugging and diagnostic tools, an improved ‘ceph df’, and some OSD backend refactoring and cleanup.

## Notable Changes

---

- ceph-brag: add tox tests (Alfredo Deza)
- common: perfcounters now use atomics and go faster (Sage Weil)
- doc: CRUSH updates (John Wilkins)
- doc: osd primary affinity (John Wilkins)
- doc: pool quotas (John Wilkins)
- doc: pre-flight doc improvements (Kevin Dalley)
- doc: switch to an unencumbered font (Ross Turk)
- doc: update openstack docs (Josh Durgin)
- fix hppa arch build (Dmitry Smirnov)
- init-ceph: continue starting other daemons on crush or mount failure (#8343, Sage Weil)
- keyvaluestore: fix hint crash (#8381, Haomai Wang)

- libcephfs-java: build against older JNI headers (Greg Farnum)
- librados: fix rados\_pool\_list bounds checks (Sage Weil)
- mds: cephfs-journal-tool (John Spray)
- mds: improve Journaler on-disk format (John Spray)
- mds, libcephfs: use client timestamp for mtime/ctime (Sage Weil)
- mds: misc encoding improvements (John Spray)
- mds: misc fixes for multi-mds (Yan, Zheng)
- mds: OPTracker integration, dump\_ops\_in\_flight (Greg Farnum)
- misc cleanup (Christophe Courtaut)
- mon: fix default replication pool ruleset choice (#8373, John Spray)
- mon: fix set cache\_target\_full\_ratio (#8440, Geoffrey Hartz)
- mon: include per-pool 'max avail' in df output (Sage Weil)
- mon: prevent EC pools from being used with cephfs (Joao Eduardo Luis)
- mon: restore original weight when auto-marked out OSDs restart (Sage Weil)
- mon: use msg header tid for MMonGetVersionReply (Ilya Dryomov)
- osd: fix bogus assert during OSD shutdown (Sage Weil)
- osd: fix clone deletion case (#8334, Sam Just)
- osd: fix filestore removal corner case (#8332, Sam Just)
- osd: fix hang waiting for osdmap (#8338, Greg Farnum)
- osd: fix interval check corner case during peering (#8104, Sam Just)
- osd: fix journal-less operation (Sage Weil)
- osd: include backend information in metadata reported to mon (Sage Weil)
- rest-api: fix help (Ailing Zhang)
- rgw: check entity permission for put\_metadata (#8428, Yehuda Sadeh)

## v0.81

---

This is the first development release since Firefly. It includes a lot of work that we delayed merging while stabilizing things. Lots of new functionality, as well as

several fixes that are baking a bit before getting backported.

## Upgrading

- CephFS support for the legacy anchor table has finally been removed. Users with file systems created before firefly should ensure that inodes with multiple hard links are modified *prior* to the upgrade to ensure that the backtraces are written properly. For example:

```
1. sudo find /mnt/cephfs -type f -links +1 -exec touch \{\}\ \;
```

- Disallow nonsensical ‘tier cache-mode’ transitions. From this point onward, ‘writeback’ can only transition to ‘forward’ and ‘forward’ can transition to 1) ‘writeback’ if there are dirty objects, or 2) any if there are no dirty objects.

## Notable Changes

- bash completion improvements (Wido den Hollander)
- brag: fixes, improvements (Loic Dachary)
- ceph-disk: handle corrupt volumes (Stuart Longlang)
- ceph-disk: partprobe as needed (Eric Eastman)
- ceph-fuse, libcephfs: asok hooks for handling session resets, timeouts (Yan, Zheng)
- ceph-fuse, libcephfs: improve traceless reply handling (Sage Weil)
- clang build fixes (John Spray, Danny Al-Gaaf)
- config: support G, M, K, etc. suffixes (Joao Eduardo Luis)
- coverity cleanups (Danny Al-Gaaf)
- doc: cache tiering (John Wilkins)
- doc: keystone integration docs (John Wilkins)
- doc: updated simple configuration guides (John Wilkins)
- libcephfs-java: fix gcj-jdk build (Dmitry Smirnov)
- librbd: check error code on cache invalidate (Josh Durgin)
- librbd: new libkrbd library for kernel map/unmap/showmapped (Ilya Dryomov)
- Makefile: fix out of source builds (Stefan Eilemann)

- mds: multi-mds fixes (Yan, Zheng)
- mds: remove legacy anchor table (Yan, Zheng)
- mds: remove legacy discover ino (Yan, Zheng)
- monclient: fix hang (Sage Weil)
- mon: prevent nonsensical cache-mode transitions (Joao Eduardo Luis)
- msgr: avoid big lock when sending (most) messages (Greg Farnum)
- osd: bound osdmap epoch skew between PGs (Sage Weil)
- osd: cache tier flushing fixes for snapped objects (Samuel Just)
- osd: fix agent early finish looping (David Zafman)
- osd: fix flush vs OpContext (Samuel Just)
- osd: fix MarkMeDown and other shutdown races (Sage Weil)
- osd: fix scrub vs cache bugs (Samuel Just)
- osd: fix trim of hitsets (Sage Weil)
- osd, msgr: fast-dispatch of OSD ops (Greg Farnum, Samuel Just)
- osd, objecter: resend ops on last\_force\_op\_resend barrier; fix cache overlay op ordering (Sage Weil)
- osd: remove obsolete classic scrub code (David Zafman)
- osd: scrub PGs with invalid stats (Sage Weil)
- osd: simple snap trimmer throttle (Sage Weil)
- osd: use FIEMAP to inform copy\_range (Haomai Wang)
- rbd-fuse: allow exposing single image (Stephen Taylor)
- rbd-fuse: fix unlink (Josh Durgin)
- removed mkcephfs (deprecated since dumpling)
- rgw: bucket link uses instance id (Yehuda Sadeh)
- rgw: fix memory leak following chunk read error (Yehuda Sadeh)
- rgw: fix URL escaping (Yehuda Sadeh)
- rgw: fix user manifest (Yehuda Sadeh)
- rgw: object and bucket rewrite functions to allow restriping old objects (Yehuda

Sadeh)

- rgw: prevent multiobject PUT race (Yehuda Sadeh)
- rgw: send user manifest header (Yehuda Sadeh)
- test\_librbd\_fsx: test krbd as well as librbd (Ilya Dryomov)

# v0.80.11 Firefly

This is a bugfix release for Firefly. This Firefly 0.80.x is nearing its planned end of life in January 2016 it may also be the last.

We recommend that all Firefly users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

- build/ops: /etc/init.d/radosgw restart does not work correctly ([issue#11140](#), [pr#5831](#), Dmitry Yatsushkevich)
- build/ops: Fix -Wno-format and -Werror=format-security options clash ([issue#13417](#), [pr#6207](#), Boris Ranto)
- build/ops: ceph-common needs python-argparse on older distros, but doesn't require it ([issue#12034](#), [pr#5217](#), Nathan Cutler)
- build/ops: ceph.spec.in running fdupes unnecessarily ([issue#12301](#), [pr#5224](#), Nathan Cutler)
- build/ops: ceph.spec.in: 50-rbd.rules conditional is wrong ([issue#12166](#), [pr#5225](#), Nathan Cutler)
- build/ops: ceph.spec.in: useless %py\_requires breaks SLE11-SP3 build ([issue#12351](#), [pr#5394](#), Nathan Cutler)
- build/ops: fedora21 has junit, not junit4 ([issue#10728](#), [pr#6203](#), Ken Dreyer, Loic Dachary)
- build/ops: upstart: configuration is too generous on restarts ([issue#11798](#), [pr#5992](#), Sage Weil)
- common: Client admin socket leaks file descriptors ([issue#11535](#), [pr#4633](#), Jon Bernard)
- common: FileStore calls syncfs(2) even it is not supported ([issue#12512](#), [pr#5529](#), Danny Al-Gaaf, Kefu Chai, Jianpeng Ma)
- common: HeartBeat: include types ([issue#13088](#), [pr#6038](#), Sage Weil)
- common: Malformed JSON command output when non-ASCII strings are present ([issue#7387](#), [pr#4635](#), Kefu Chai, Tim Serong)
- common: Memory leak in Mutex.cc, pthread\_mutexattr\_init without pthread\_mutexattr\_destroy ([issue#11762](#), [pr#5403](#), Ketor Meng)

- common: Thread:pthread\_attr\_destroy(thread\_attr) when done with it ([issue#12570](#), [pr#6325](#), Piotr Dałek, Zheng Qiankun)
- common: ThreadPool add/remove work queue methods not thread safe ([issue#12662](#), [pr#5991](#), Jason Dillaman)
- common: buffer: critical bufferlist::zero bug ([issue#12252](#), [pr#5388](#), Haomai Wang)
- common: log: take mutex while opening fd ([issue#12465](#), [pr#5406](#), Samuel Just)
- common: recursive lock of md\_config\_t (0) ([issue#12614](#), [pr#5814](#), Josh Durgin)
- crush: take crashes due to invalid arg ([issue#11602](#), [pr#4769](#), Sage Weil)
- doc: backport v0.80.10 release notes to firefly ([issue#11090](#), [pr#5307](#), Loic Dachary, Sage Weil)
- doc: update docs to point to download.ceph.com ([issue#13162](#), [pr#5993](#), Alfredo Deza)
- fs: MDSMonitor: handle MDSBeacon messages properly ([issue#11590](#), [pr#5199](#), Kefu Chai)
- fs: client nonce collision due to unshared pid namespaces ([issue#13032](#), [pr#6087](#), Josh Durgin, Sage Weil)
- librbd: Objectcacher setting max object counts too low ([issue#7385](#), [pr#4639](#), Jason Dillaman)
- librbd: aio calls may block ([issue#11056](#), [pr#4854](#), Haomai Wang, Sage Weil, Jason Dillaman)
- librbd: internal.cc: 1967: FAILED assert(watchers.size() == 1) ([issue#12176](#), [pr#5171](#), Jason Dillaman)
- mon: Clock skew causes missing summary and confuses Calamari ([issue#11877](#), [pr#4867](#), Thorsten Behrens)
- mon: EC pools are not allowed as cache pools, disallow in the mon ([issue#11650](#), [pr#5389](#), Samuel Just)
- mon: Make it more difficult to delete pools in firefly ([issue#11800](#), [pr#4788](#), Sage Weil)
- mon: MonitorDBStore: get\_next\_key() only if prefix matches ([issue#11786](#), [pr#5360](#), Joao Eduardo Luis)
- mon: PaxosService: call post\_refresh() instead of post\_paxos\_update() ([issue#11470](#), [pr#5358](#), Joao Eduardo Luis)
- mon: add a cache layer over MonitorDBStore ([issue#12638](#), [pr#5698](#), Kefu Chai)

- mon: adding existing pool as tier with -force-nonempty clobbers removed\_snaps ([issue#11493](#), [pr#5236](#), Sage Weil, Samuel Just)
- mon: ceph fails to compile with boost 1.58 ([issue#11576](#), [pr#5129](#), Kefu Chai)
- mon: does not check for IO errors on every transaction ([issue#13089](#), [pr#6091](#), Sage Weil)
- mon: get pools health'info have error ([issue#12402](#), [pr#5410](#), renhwztetecs)
- mon: increase globalid default for firefly ([issue#13255](#), [pr#6010](#), Sage Weil)
- mon: pgmonitor: wrong at/near target max" reporting ([issue#12401](#), [pr#5409](#), huangjun)
- mon: register\_new\_pgs() should check ruleno instead of its index ([issue#12210](#), [pr#5404](#), Xinze Chi)
- mon: scrub error (osdmap encoding mismatch?) upgrading from 0.80 to ~0.80.2 ([issue#8815](#), [issue#8674](#), [issue#9064](#), [pr#5200](#), Sage Weil, Zhiqiang Wang, Samuel Just)
- mon: the output is wrong when running ceph osd reweight ([issue#12251](#), [pr#5408](#), Joao Eduardo Luis)
- objecter: can get stuck in redirect loop if osdmap epoch == last\_force\_op\_resend ([issue#11026](#), [pr#4597](#), Jianpeng Ma, Sage Weil)
- objecter: pg listing can deadlock when throttling is in use ([issue#9008](#), [pr#5043](#), Guang Yang)
- objecter: resend linger ops on split ([issue#9806](#), [pr#5062](#), Josh Durgin, Samuel Just)
- osd: Cleanup boost optionals for boost 1.56 ([issue#9983](#), [pr#5039](#), William A. Kennington III)
- osd: LibRadosTwoPools[EC]PP.PromoteSnap failure ([issue#10052](#), [pr#5050](#), Sage Weil)
- osd: Mutex Assert from PipeConnection::try\_get\_pipe ([issue#12437](#), [pr#5815](#), David Zafman)
- osd: PG stuck with remapped ([issue#9614](#), [pr#5044](#), Guang Yang)
- osd: PG::handle\_advance\_map: on\_pool\_change after handling the map change ([issue#12809](#), [pr#5988](#), Samuel Just)
- osd: PGLog: split divergent priors as well ([issue#11069](#), [pr#4631](#), Samuel Just)
- osd: PGLog::proc\_replica\_log: correctly handle case where entries between olog.head and log.tail were split out ([issue#11358](#), [pr#5287](#), Samuel Just)

- osd: WBThrottle::clear\_object: signal on cond when we reduce throttle values ([issue#12223](#), [pr#5822](#), Samuel Just)
- osd: cache full mode still skips young objects ([issue#10006](#), [pr#5051](#), Xinze Chi, Zhiqiang Wang)
- osd: crash creating/deleting pools ([issue#12429](#), [pr#5526](#), John Spray)
- osd: explicitly specify OSD features in MOSDBoot ([issue#10911](#), [pr#4960](#), Sage Weil)
- osd: is\_new\_interval() fixes ([issue#11771](#), [issue#10399](#), [pr#5726](#), Samuel Just, Jason Dillaman)
- osd: make the all osd/filestore thread pool suicide timeouts separately configurable ([issue#11439](#), [pr#5823](#), Samuel Just)
- osd: object creation by write cannot use an offset on an erasure coded pool ([issue#11507](#), [pr#4632](#), Jianpeng Ma, Loic Dachary)
- osd: os/FileJournal: Fix journal write fail, align for direct io ([issue#12943](#), [pr#5619](#), Xie Rui)
- osd: os/PGLog.cc: 732: FAILED assert(log.log.size() == log\_keys\_debug.size()) ([issue#12652](#), [pr#5820](#), Sage Weil)
- osd: read on chunk-aligned xattr not handled ([issue#12309](#), [pr#5235](#), Sage Weil)
- rgw: Change variable length array of std::strings (not legal in C++) to std::vector<std::string> ([issue#12467](#), [pr#4583](#), Daniel J. Hofmann)
- rgw: Civetweb RGW appears to report full size of object as downloaded when only partially downloaded ([issue#11851](#), [pr#5234](#), Yehuda Sadeh)
- rgw: Keystone PKI token expiration is not enforced ([issue#11367](#), [pr#4765](#), Anton Aksola)
- rgw: Object copy bug ([issue#11639](#), [pr#4762](#), Javier M. Mellid)
- rgw: RGW returns requested bucket name raw in "Bucket" response header ([issue#11860](#), [issue#12537](#), [pr#5730](#), Yehuda Sadeh, Wido den Hollander)
- rgw: Swift API: response for PUT on /container does not contain the mandatory Content-Length header when FCGI is used ([issue#11036](#), [pr#5170](#), Radoslaw Zarzynski)
- rgw: content length parsing calls `strtol()` instead of `strtoll()` ([issue#10701](#), [pr#5997](#), Yehuda Sadeh)
- rgw: delete bucket does not remove .bucket.meta file ([issue#11149](#), [pr#4641](#), Orit Wasserman)

- rgw: doesn't return 'x-timestamp' in header which is used by 'View Details' of OpenStack ([issue#8911](#), [pr#4584](#), Yehuda Sadeh)
- rgw: init some manifest fields when handling explicit objs ([issue#11455](#), [pr#5729](#), Yehuda Sadeh)
- rgw: logfile does not get chowned properly ([issue#12073](#), [pr#5233](#), Thorsten Behrens)
- rgw: logrotate.conf calls service with wrong init script name ([issue#12043](#), [pr#5390](#), wuxingyi)
- rgw: quota not respected in POST object ([issue#11323](#), [pr#4642](#), Sergey Arkhipov)
- rgw: swift smoke test fails on TestAccountUTF8 ([issue#11091](#), [issue#11438](#), [issue#12939](#), [issue#12157](#), [issue#12158](#), [issue#12363](#), [pr#5532](#), Radoslaw Zarzynski, Orit Wasserman, Robin H. Johnson)
- rgw: use correct objv\_tracker for bucket instance ([issue#11416](#), [pr#4535](#), Yehuda Sadeh)
- tests: ceph-fuse crash in test\_client\_recovery ([issue#12673](#), [pr#5813](#), Loic Dachary)
- tests: kernel\_untar\_build fails on EL7 ([issue#11758](#), [pr#6000](#), Greg Farnum)
- tests: qemu workunit refers to apt-mirror.front.sepia.ceph.com ([issue#13420](#), [pr#6328](#), Yuan Zhou, Sage Weil)
- tools: src/ceph-disk : disk zap sgdisk invocation ([issue#11143](#), [pr#4636](#), Thorsten Behrens, Owen Synge)
- tools: ceph-disk: sometimes the journal symlink is not created ([issue#10146](#), [pr#5541](#), Dan van der Ster)
- tools: ceph-disk: support NVMe device partitions ([issue#11612](#), [pr#4771](#), Ilja Slepnev)
- tools: ceph-post-file fails on rhel7 ([issue#11836](#), [pr#5037](#), Joseph McDonald, Sage Weil)
- tools: ceph\_argparse\_flag has no regular 3rd parameter ([issue#11543](#), [pr#4582](#), Thorsten Behrens)
- tools: use a new disk as journal disk,ceph-disk prepare fail ([issue#10983](#), [pr#4630](#), Loic Dachary)

## v0.80.10 Firefly

This is a bugfix release for Firefly.

We recommend that all Firefly users upgrade.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build/ops: ceph.spec.in: package mkcephfs on EL6 ([issue#11955](#), [pr#4924](#), Ken Dreyer)
- build/ops: debian: ceph-test and rest-bench debug packages should require their respective binary packages ([issue#11673](#), [pr#4766](#), Ken Dreyer)
- build/ops: run RGW as root ([issue#11453](#), [pr#4638](#), Ken Dreyer)
- common: messages/MWatchNotify: include an error code in the message ([issue#9193](#), [pr#3944](#), Sage Weil)
- common: Rados.shutdown() dies with Illegal instruction (core dumped) ([issue#10153](#), [pr#3963](#), Federico Simoncelli)
- common: SimpleMessenger: allow RESETSESSION whenever we forget an endpoint ([issue#10080](#), [pr#3915](#), Greg Farnum)
- common: WorkQueue: make wait timeout on empty queue configurable ([issue#10817](#), [pr#3941](#), Samuel Just)
- crush: set\_choose\_tries = 100 for erasure code rulesets ([issue#10353](#), [pr#3824](#), Loic Dachary)
- doc: backport ceph-disk man page to Firefly ([issue#10724](#), [pr#3936](#), Nilamdyuti Goswami)
- doc: Fix ceph command manpage to match ceph -h ([issue#10676](#), [pr#3996](#), David Zafman)
- fs: mount.ceph: avoid spurious error message ([issue#10351](#), [pr#3927](#), Yan, Zheng)
- librados: Fix memory leak in python rados bindings ([issue#10723](#), [pr#3935](#), Josh Durgin)
- librados: fix resources leakage in RadosClient::connect() ([issue#10425](#), [pr#3828](#), Radoslaw Zarzynski)
- librados: Translate operation flags from C APIs ([issue#10497](#), [pr#3930](#), Matt Richards)
- librbd: acquire cache\_lock before refreshing parent ([issue#5488](#), [pr#4206](#), Jason Dillaman)
- librbd: snap\_remove should ignore -ENOENT errors ([issue#11113](#), [pr#4245](#), Jason

Dillaman)

- mds: fix assertion caused by system clock backwards ([issue#11053](#), [pr#3970](#), Yan, Zheng)
- mon: ignore osd failures from before up\_from ([issue#10762](#), [pr#3937](#), Sage Weil)
- mon: MonCap: take EntityName instead when expanding profiles ([issue#10844](#), [pr#3942](#), Joao Eduardo Luis)
- mon: Monitor: fix timecheck rounds period ([issue#10546](#), [pr#3932](#), Joao Eduardo Luis)
- mon: OSDMonitor: do not trust small values in osd epoch cache ([issue#10787](#), [pr#3823](#), Sage Weil)
- mon: OSDMonitor: fallback to json-pretty in case of invalid formatter ([issue#9538](#), [pr#4475](#), Loic Dachary)
- mon: PGMonitor: several stats output error fixes ([issue#10257](#), [pr#3826](#), Joao Eduardo Luis)
- objecter: fix map skipping ([issue#9986](#), [pr#3952](#), Ding Dinghua)
- osd: cache tiering: fix the atime logic of the eviction ([issue#9915](#), [pr#3949](#), Zhiqiang Wang)
- osd: cancel\_pull: requeue waiters ([issue#11244](#), [pr#4415](#), Samuel Just)
- osd: check that source OSD is valid for MOSDRepScrub ([issue#9555](#), [pr#3947](#), Sage Weil)
- osd: DBObjectMap: lock header\_lock on sync() ([issue#9891](#), [pr#3948](#), Samuel Just)
- osd: do not ignore deleted pgs on startup ([issue#10617](#), [pr#3933](#), Sage Weil)
- osd: ENOENT on clone ([issue#11199](#), [pr#4385](#), Samuel Just)
- osd: erasure-code-profile set races with erasure-code-profile rm ([issue#11144](#), [pr#4383](#), Loic Dachary)
- osd: FAILED assert(soid < scrubber.start || soid >= scrubber.end) ([issue#11156](#), [pr#4185](#), Samuel Just)
- osd: FileJournal: fix journalq population in do\_read\_entry() ([issue#6003](#), [pr#3960](#), Samuel Just)
- osd: fix negative degraded objects during backfilling ([issue#7737](#), [pr#4021](#), Guang Yang)
- osd: get the currently atime of the object in cache pool for eviction ([issue#9985](#), [pr#3950](#), Sage Weil)

- osd: load\_pgs: we need to handle the case where an upgrade from earlier versions which ignored non-existent pgs resurrects a pg with a prehistoric osdmap ([issue#11429](#), [pr#4556](#), Samuel Just)
- osd: ObjectStore: Don't use largest\_data\_off to calc data\_align. ([issue#10014](#), [pr#3954](#), Jianpeng Ma)
- osd: osd\_types: op\_queue\_age\_hist and fs\_perf\_stat should be in osd\_stat\_t::o... ([issue#10259](#), [pr#3827](#), Samuel Just)
- osd: PG::actingset should be used when checking the number of acting OSDs for... ([issue#11454](#), [pr#4453](#), Guang Yang)
- osd: PG::all\_unfound\_are\_queried\_or\_lost for non-existent osds ([issue#10976](#), [pr#4416](#), Mykola Golub)
- osd: PG: always clear\_primary\_state ([issue#10059](#), [pr#3955](#), Samuel Just)
- osd: PGLog.h: 279: FAILED assert(log.log.size() == log\_keys\_debug.size()) ([issue#10718](#), [pr#4382](#), Samuel Just)
- osd: PGLog: include rollback\_info\_trimmed\_to in (read|write)\_log ([issue#10157](#), [pr#3964](#), Samuel Just)
- osd: pg stuck stale after create with activation delay ([issue#11197](#), [pr#4384](#), Samuel Just)
- osd: ReplicatedPG: fail a non-blocking flush if the object is being scrubbed ([issue#8011](#), [pr#3943](#), Samuel Just)
- osd: ReplicatedPG::on\_change: clean up callbacks\_for\_degraded\_object ([issue#8753](#), [pr#3940](#), Samuel Just)
- osd: ReplicatedPG::scan\_range: an object can disappear between the list and t... ([issue#10150](#), [pr#3962](#), Samuel Just)
- osd: requeue blocked op before flush it was blocked on ([issue#10512](#), [pr#3931](#), Sage Weil)
- rgw: check for timestamp for s3 keystone auth ([issue#10062](#), [pr#3958](#), Abhishek Lekshmanan)
- rgw: civetweb should use unique request id ([issue#11720](#), [pr#4780](#), Orit Wasserman)
- rgw: don't allow negative / invalid content length ([issue#11890](#), [pr#4829](#), Yehuda Sadeh)
- rgw: fail s3 POST auth if keystone not configured ([issue#10698](#), [pr#3966](#), Yehuda Sadeh)
- rgw: flush xml header on get acl request ([issue#10106](#), [pr#3961](#), Yehuda Sadeh)

- rgw: generate new tag for object when setting object attrs ([issue#11256](#), [pr#4571](#), Yehuda Sadeh)
- rgw: generate the “Date” HTTP header for civetweb. ([issue#11871](#), [11891](#), [pr#4851](#), Radoslaw Zarzynski)
- rgw: keystone token cache does not work correctly ([issue#11125](#), [pr#4414](#), Yehuda Sadeh)
- rgw: merge manifests correctly when there’s prefix override ([issue#11622](#), [pr#4697](#), Yehuda Sadeh)
- rgw: send appropriate op to cancel bucket index pending operation ([issue#10770](#), [pr#3938](#), Yehuda Sadeh)
- rgw: shouldn’t need to disable rgw\_socket\_path if frontend is configured ([issue#11160](#), [pr#4275](#), Yehuda Sadeh)
- rgw: Swift API. Dump container’s custom metadata. ([issue#10665](#), [pr#3934](#), Dmytro Iurchenko)
- rgw: Swift API. Support for X-Remove-Container-Meta-{key} header. ([issue#10475](#), [pr#3929](#), Dmytro Iurchenko)
- rgw: use correct objv\_tracker for bucket instance ([issue#11416](#), [pr#4379](#), Yehuda Sadeh)
- tests: force checkout of submodules ([issue#11157](#), [pr#4079](#), Loic Dachary)
- tools: Backport ceph-objectstore-tool changes to firefly ([issue#12327](#), [pr#3866](#), David Zafman)
- tools: ceph-objectstore-tool: Output only unsupported features when incompatible ([issue#11176](#), [pr#4126](#), David Zafman)
- tools: ceph-objectstore-tool: Use exit status 11 for incompatible import attempt... ([issue#11139](#), [pr#4129](#), David Zafman)
- tools: Fix do\_autogen.sh so that -L is allowed ([issue#11303](#), [pr#4247](#), Alfredo Deza)

## v0.80.9 Firefly

---

This is a bugfix release for firefly. It fixes a performance regression in librbd, an important CRUSH misbehavior (see below), and several RGW bugs. We have also backported support for flock/fcntl locks to ceph-fuse and libcephfs.

We recommend that all Firefly users upgrade.

For more detailed information, see [the complete changelog](#).

# Adjusting CRUSH maps

- This point release fixes several issues with CRUSH that trigger excessive data migration when adjusting OSD weights. These are most obvious when a very small weight change (e.g., a change from 0 to .01) triggers a large amount of movement, but the same set of bugs can also lead to excessive (though less noticeable) movement in other cases.

However, because the bug may already have affected your cluster, fixing it may trigger movement *back* to the more correct location. For this reason, you must manually opt-in to the fixed behavior.

In order to set the new tunable to correct the behavior:

```
1. ceph osd crush set-tunable straw_calc_version 1
```

Note that this change will have no immediate effect. However, from this point forward, any ‘straw’ bucket in your CRUSH map that is adjusted will get non-buggy internal weights, and that transition may trigger some rebalancing.

You can estimate how much rebalancing will eventually be necessary on your cluster with:

```
1. ceph osd getcrushmap -o /tmp/cm
2. crushtool -i /tmp/cm --num-rep 3 --test --show-mappings > /tmp/a 2>&1
3. crushtool -i /tmp/cm --set-straw-calc-version 1 -o /tmp/cm2
4. crushtool -i /tmp/cm2 --reweight -o /tmp/cm2
5. crushtool -i /tmp/cm2 --num-rep 3 --test --show-mappings > /tmp/b 2>&1
6. wc -l /tmp/a          # num total mappings
7. diff -u /tmp/a /tmp/b | grep -c ^+  # num changed mappings
8.
9. Divide the number of changed lines by the total number of lines in
10. /tmp/a. We've found that most clusters are under 10%.
11.
12. You can force all of this rebalancing to happen at once with::
13.
14. ceph osd crush reweight-all
15.
16. Otherwise, it will happen at some unknown point in the future when
17. CRUSH weights are next adjusted.
```

## Notable Changes

- ceph-fuse: flock, fcntl lock support (Yan, Zheng, Greg Farnum)
- crush: fix straw bucket weight calculation, add straw\_calc\_version tunable (#10095 Sage Weil)

- crush: fix tree bucket (Rongzu Zhu)
- crush: fix underflow of tree weights (Loic Dachary, Sage Weil)
- crushtool: add -reweight (Sage Weil)
- librbd: complete pending operations before losing image (#10299 Jason Dillaman)
- librbd: fix read caching performance regression (#9854 Jason Dillaman)
- librbd: gracefully handle deleted/renamed pools (#10270 Jason Dillaman)
- mon: fix dump of chooseleaf\_vary\_r tunable (Sage Weil)
- osd: fix PG ref leak in snaptrimmer on peering (#10421 Kefu Chai)
- osd: handle no-op write with snapshot (#10262 Sage Weil)
- radosgw-admin: create subuser when creating user (#10103 Yehuda Sadeh)
- rgw: change multipart uplaod id magic (#10271 Georgio Dimitrakakis, Yehuda Sadeh)
- rgw: don't overwrite bucket/object owner when setting ACLs (#10978 Yehuda Sadeh)
- rgw: enable IPv6 for embedded civetweb (#10965 Yehuda Sadeh)
- rgw: fix partial swift GET (#10553 Yehuda Sadeh)
- rgw: fix quota disable (#9907 Dong Lei)
- rgw: index swift keys appropriately (#10471 Hemant Burman, Yehuda Sadeh)
- rgw: make setattrs update bucket index (#5595 Yehuda Sadeh)
- rgw: pass civetweb configurables (#10907 Yehuda Sadeh)
- rgw: remove swift user manifest (DLO) hash calculation (#9973 Yehuda Sadeh)
- rgw: return correct len for 0-len objects (#9877 Yehuda Sadeh)
- rgw: S3 object copy content-type fix (#9478 Yehuda Sadeh)
- rgw: send ETag on S3 object copy (#9479 Yehuda Sadeh)
- rgw: send HTTP status reason explicitly in fastcgi (Yehuda Sadeh)
- rgw: set ulimit -n from sysvinit (el6) init script (#9587 Sage Weil)
- rgw: update swift subuser permission masks when authenticating (#9918 Yehuda Sadeh)
- rgw: URL decode query params correctly (#10271 Georgio Dimitrakakis, Yehuda Sadeh)

- rgw: use attrs when reading object attrs (#10307 Yehuda Sadeh)
- rgw: use rn for http headers (#9254 Benedikt Fraunhofer, Yehuda Sadeh)

## v0.80.8 Firefly

---

This is a long-awaited bugfix release for firefly. It has several important (but relatively rare) OSD peering fixes, performance issues when snapshots are trimmed, several RGW fixes, a paxos corner case fix, and some packaging updates.

We recommend that all users for v0.80.x firefly upgrade when it is convenient to do so.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build: remove stack-execute bit from assembled code sections (#10114 Dan Mick)
- ceph-disk: fix dmcrypt key permissions (#9785 Loic Dachary)
- ceph-disk: fix keyring location (#9653 Loic Dachary)
- ceph-disk: make partition checks more robust (#9721 #9665 Loic Dachary)
- ceph: cleanly shut down librados context on shutdown (#8797 Dan Mick)
- common: add \$cctid config metavariable (#6228 Adam Crume)
- crush: align rule and ruleset ids (#9675 Xiaoxi Chen)
- crush: fix negative weight bug during create\_or\_move\_item (#9998 Paweł Sadowski)
- crush: fix potential buffer overflow in erasure rules (#9492 Johnu George)
- debian: fix python-ceph -> ceph file movement (Sage Weil)
- libcephfs,ceph-fuse: fix flush tid wraparound bug (#9869 Greg Farnum, Yan, Zheng)
- libcephfs: close fd before umount (#10415 Yan, Zheng)
- librados: fix crash from C API when read timeout is enabled (#9582 Sage Weil)
- librados: handle reply race with pool deletion (#10372 Sage Weil)
- librbd: cap memory utilization for read requests (Jason Dillaman)
- librbd: do not close a closed parent image on failure (#10030 Jason Dillaman)
- librbd: fix diff tests (#10002 Josh Durgin)

- librbd: protect list\_children from invalid pools (#10123 Jason Dillaman)
- make check improvemens (Loic Dachary)
- mds: fix ctime updates (#9514 Greg Farnum)
- mds: fix journal import tool (#10025 John Spray)
- mds: fix rare NULL deref in cap flush handler (Greg Farnum)
- mds: handle unknown lock messages (Yan, Zheng)
- mds: store backtrace for straydir (Yan, Zheng)
- mon: abort startup if disk is full (#9502 Joao Eduardo Luis)
- mon: add paxos instrumentation (Sage Weil)
- mon: fix double-free in rare OSD startup path (Sage Weil)
- mon: fix osdmap trimming (#9987 Sage Weil)
- mon: fix paxos corner cases (#9301 #9053 Sage Weil)
- osd: cancel callback on blacklisted watchers (#8315 Samuel Just)
- osd: cleanly abort set-alloc-hint operations during upgrade (#9419 David Zafman)
- osd: clear rollback PG metadata on PG deletion (#9293 Samuel Just)
- osd: do not abort deep scrub if hinfo is missing (#10018 Loic Dachary)
- osd: erasure-code regression tests (Loic Dachary)
- osd: fix distro metadata reporting for SUSE (#8654 Danny Al-Gaaf)
- osd: fix full OSD checks during backfill (#9574 Samuel Just)
- osd: fix ioprio parsing (#9677 Loic Dachary)
- osd: fix journal direct-io shutdown (#9073 Mark Kirkwood, Ma Jianpeng, Somnath Roy)
- osd: fix journal dump (Ma Jianpeng)
- osd: fix occasional stall during peering or activation (Sage Weil)
- osd: fix past\_interval display bug (#9752 Loic Dachary)
- osd: fix rare crash triggered by admin socket dump\_ops\_in\_filght (#9916 Dong Lei)
- osd: fix snap trimming performance issues (#9487 #9113 Samuel Just, Sage Weil, Dan van der Ster, Florian Haas)

- osd: fix snapdir handling on cache eviction (#8629 Sage Weil)
- osd: handle map gaps in map advance code (Sage Weil)
- osd: handle undefined CRUSH results in interval check (#9718 Samuel Just)
- osd: include shard in JSON dump of ghobject (#10063 Loic Dachary)
- osd: make backfill reservation denial handling more robust (#9626 Samuel Just)
- osd: make misdirected op checks handle EC + primary affinity (#9835 Samuel Just, Sage Weil)
- osd: mount XFS with inode64 by default (Sage Weil)
- osd: other misc bugs (#9821 #9875 Samuel Just)
- rgw: add .log to default log path (#9353 Alexandre Marangone)
- rgw: clean up fcgi request context (#10194 Yehuda Sadeh)
- rgw: convert header underscores to dashes (#9206 Yehuda Sadeh)
- rgw: copy object data if copy target is in different pool (#9039 Yehuda Sadeh)
- rgw: don't try to authenticate CORS preflight request (#8718 Robert Hubbard, Yehuda Sadeh)
- rgw: fix civetweb URL decoding (#8621 Yehuda Sadeh)
- rgw: fix hash calculation during PUT (Yehuda Sadeh)
- rgw: fix misc bugs (#9089 #9201 Yehuda Sadeh)
- rgw: fix object tail test (#9226 Sylvain Munaut, Yehuda Sadeh)
- rgw: make sysvinit script run rgw under systemd context as needed (#10125 Loic Dachary)
- rgw: separate civetweb log from rgw log (Yehuda Sadeh)
- rgw: set length for keystone token validations (#7796 Mark Kirkwood, Yehuda Sadeh)
- rgw: subuser creation fixes (#8587 Yehuda Sadeh)
- rpm: misc packaging improvements (Sandon Van Ness, Dan Mick, Erik Logthenberg, Boris Ranto)
- rpm: use standard udev rules for CentOS7/RHEL7 (#9747 Loic Dachary)

## v0.80.7 Firefly

This release fixes a few critical issues with v0.80.6, particularly with clusters running mixed versions.

We recommend that all v0.80.x Firefly users upgrade to this release.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- osd: fix invalid memory reference in log trimming (#9731 Samuel Just)
- osd: fix use-after-free in cache tiering code (#7588 Sage Weil)
- osd: remove bad backfill assertion for mixed-version clusters (#9696 Samuel Just)

## v0.80.6 Firefly

---

This is a major bugfix release for firefly, fixing a range of issues in the OSD and monitor, particularly with cache tiering. There are also important fixes in librados, with the watch/notify mechanism used by librbd, and in radosgw.

A few pieces of new functionality have been backported, including improved ‘ceph df’ output (view amount of writeable space per pool), support for non-default cluster names when using sysvinit or systemd, and improved (and fixed) support for dmcrypt.

We recommend that all v0.80.x Firefly users upgrade to this release.

For more detailed information, see [the complete changelog](#).

## Notable Changes

---

- build: fix atomic64\_t on i386 (#8969 Sage Weil)
- build: fix build on alpha (Michael Cree, Dmitry Smirnov)
- build: fix build on\_hppa (Dmitry Smirnov)
- build: fix yasm detection on x32 arch (Sage Weil)
- ceph-disk: fix ‘list’ function with dmcrypt (Sage Weil)
- ceph-disk: fix dmcrypt support (Alfredo Deza)
- ceph: allow non-default cluster to be specified (#8944)
- common: fix dup log messages to mon (#9080 Sage Weil)
- global: write pid file when -f is used (systemd, upstart) (Alexandre Oliva)

- librados: fix crash when read timeout is enabled (#9362 Matthias Kiefer, Sage Weil)
- librados: fix lock leaks in error paths (#9022 Pavan Rallabhandi)
- librados: fix watch resend on PG acting set change (#9220 Samuel Just)
- librados: python: fix aio\_read handling with 0 (Mohammad Salehe)
- librbd: add interface to invalidate cached data (Josh Durgin)
- librbd: fix crash when using clone of flattened image (#8845 Josh Durgin)
- librbd: fix error path cleanup on open (#8912 Josh Durgin)
- librbd: fix null pointer check (Danny Al-Gaaf)
- librbd: limit dirty object count (Haomai Wang)
- mds: fix rstats for root and mdsdir (Yan, Zheng)
- mon: add 'get' command for new cache tier pool properties (Joao Eduardo Luis)
- mon: add 'osd pool get-quota' (#8523 Joao Eduardo Luis)
- mon: add cluster fingerprint (Sage Weil)
- mon: disallow nonsensical cache-mode transitions (#8155 Joao Eduardo Luis)
- mon: fix cache tier rounding error on i386 (Sage Weil)
- mon: fix occasional memory leak (#9176 Sage Weil)
- mon: fix reported latency for 'osd perf' (#9269 Samuel Just)
- mon: include 'max avail' in 'ceph df' output (Sage Weil, Xiaoxi Chen)
- mon: persistently mark pools where scrub may find incomplete clones (#8882 Sage Weil)
- mon: preload erasure plugins (Loic Dachary)
- mon: prevent cache-specific settings on non-tier pools (#8696 Joao Eduardo Luis)
- mon: reduce log spam (Aanchal Agrawal, Sage Weil)
- mon: warn when cache pools have no hit\_sets enabled (Sage Weil)
- msgr: fix trivial memory leak (Sage Weil)
- osd: automatically scrub PGs with invalid stats (#8147 Sage Weil)
- osd: avoid sharing PG metadata that is not durable (Samuel Just)

- osd: cap hit\_set size (#9339 Samuel Just)
- osd: create default erasure profile if needed (#8601 Loic Dachary)
- osd: dump tid as JSON int (not string) where appropriate (Joao Eduardo Luis)
- osd: encode blacklist in deterministic order (#9211 Sage Weil)
- osd: fix behavior when cache tier has no hit\_sets enabled (#8982 Sage Weil)
- osd: fix cache tier flushing of snapshots (#9054 Samuel Just)
- osd: fix cache tier op ordering when going from full to non-full (#8931 Sage Weil)
- osd: fix crash on dup recovery reservation (#8863 Sage Weil)
- osd: fix division by zero when pg\_num adjusted with no OSDs (#9052 Sage Weil)
- osd: fix hint crash in experimental keyvaluestore\_dev backend (Hoamai Wang)
- osd: fix leak in copyfrom cancellation (#8894 Samuel Just)
- osd: fix locking for copyfrom finish (#8889 Sage Weil)
- osd: fix long filename handling in backend (#8701 Sage Weil)
- osd: fix min\_size check with backfill (#9497 Samuel Just)
- osd: fix mount/remount sync race (#9144 Sage Weil)
- osd: fix object listing + erasure code bug (Guang Yang)
- osd: fix race on reconnect to failed OSD (#8944 Greg Farnum)
- osd: fix recovery reservation deadlock (Samuel Just)
- osd: fix tiering agent arithmetic for negative values (#9082 Karan Singh)
- osd: improve shutdown order (#9218 Sage Weil)
- osd: improve subop discard logic (#9259 Samuel Just)
- osd: introduce optional sleep, io priority for scrub and snap trim (Sage Weil)
- osd: make scrub check for and remove stale erasure-coded objects (Samuel Just)
- osd: misc fixes (#9481 #9482 #9179 Sameul Just)
- osd: mix keyvaluestore\_dev improvements (Haomai Wang)
- osd: only require CRUSH features for rules that are used (#8963 Sage Weil)
- osd: preload erasure plugins on startup (Loic Dachary)

- osd: prevent PGs from falling behind when consuming OSDMaps (#7576 Sage Weil)
- osd: prevent old clients from using tiered pools (#8714 Sage Weil)
- osd: set min\_size on erasure pools to data chunk count (Sage Weil)
- osd: trim old erasure-coded objects more aggressively (Samuel Just)
- rados: enforce erasure code alignment (Lluis Pamies-Juarez)
- rgw: align object stripes with erasure pool alignment (#8442 Yehuda Sadeh)
- rgw: don't send error body on HEAD for civetweb (#8539 Yehuda Sadeh)
- rgw: fix crash in CORS preflight request (Yehuda Sadeh)
- rgw: fix decoding of + in URL (#8702 Brian Rak)
- rgw: fix object removal on object create (#8972 Patrycja Szabowska, Yehuda Sadeh)
- systemd: use systemd-run when starting radosgw (JuanJose Galvez)
- sysvinit: support non-default cluster name (Alfredo Deza)

## v0.80.5 Firefly

---

This release fixes a few important bugs in the radosgw and fixes several packaging and environment issues, including OSD log rotation, systemd environments, and daemon restarts on upgrade.

We recommend that all v0.80.x Firefly users upgrade, particularly if they are using upstart, systemd, or radosgw.

## Notable Changes

---

- ceph-dencoder: do not needlessly link to librgw, librados, etc. (Sage Weil)
- do not needlessly link binaries to leveldb (Sage Weil)
- mon: fix mon crash when no auth keys are present (#8851, Joao Eduardo Luis)
- osd: fix cleanup (and avoid occasional crash) during shutdown (#7981, Sage Weil)
- osd: fix log rotation under upstart (Sage Weil)
- rgw: fix multipart upload when object has irregular size (#8846, Yehuda Sadeh, Sylvain Munaut)
- rgw: improve bucket listing S3 compatibility (#8858, Yehuda Sadeh)
- rgw: improve delimited bucket listing (Yehuda Sadeh)

- rpm: do not restart daemons on upgrade (#8849, Alfredo Deza)

For more detailed information, see [the complete changelog](#).

## v0.80.4 Firefly

---

This Firefly point release fixes an potential data corruption problem when ceph-osd daemons run on top of XFS and service Firefly librbd clients. A recently added allocation hint that RBD utilizes triggers an XFS bug on some kernels (Linux 3.2, and likely others) that leads to data corruption and deep-scrub errors (and inconsistent PGs). This release avoids the situation by disabling the allocation hint until we can validate which kernels are affected and/or are known to be safe to use the hint on.

We recommend that all v0.80.x Firefly users urgently upgrade, especially if they are using RBD.

## Notable Changes

---

- osd: disable XFS extsize hint by default (#8830, Samuel Just)
- rgw: fix extra data pool default name (Yehuda Sadeh)

For more detailed information, see [the complete changelog](#).

## v0.80.3 Firefly

---

This is the third Firefly point release. It includes a single fix for a radosgw regression that was discovered in v0.80.2 right after it was released.

We recommend that all v0.80.x Firefly users upgrade.

## Notable Changes

---

- radosgw: fix regression in manifest decoding (#8804, Sage Weil)

For more detailed information, see [the complete changelog](#).

## v0.80.2 Firefly

---

This is the second Firefly point release. It contains a range of important fixes, including several bugs in the OSD cache tiering, some compatibility checks that affect upgrade situations, several radosgw bugs, and an irritating and unnecessary feature bit check that prevents older clients from communicating with a cluster with any erasure coded pools.

One someone large change in this point release is that the ceph RPM package is separated into a ceph and ceph-common package, similar to Debian. The ceph-common package contains just the client libraries without any of the server-side daemons.

We recommend that all v0.80.x Firefly users skip this release and use v0.80.3.

## Notable Changes

---

- ceph-disk: better debug logging (Alfredo Deza)
- ceph-disk: fix preparation of OSDs with dmcrypt (#6700, Stephen F Taylor)
- ceph-disk: partprobe on prepare to fix dm-crypt (#6966, Eric Eastman)
- do not require ERASURE\_CODE feature from clients (#8556, Sage Weil)
- libcephfs-java: build with older JNI headers (Greg Farnum)
- libcephfs-java: fix build with gcj-jdk (Dmitry Smirnov)
- librados: fix osd op tid for redirected ops (#7588, Samuel Just)
- librados: fix rados\_pool\_list buffer bounds checks (#8447, Sage Weil)
- librados: resend ops when pool overlay changes (#8305, Sage Weil)
- librbd, ceph-fuse: reduce CPU overhead for clean object check in cache (Haomai Wang)
- mon: allow deletion of cephfs pools (John Spray)
- mon: fix default pool ruleset choice (#8373, John Spray)
- mon: fix health summary for mon low disk warning (Sage Weil)
- mon: fix ‘osd pool set <pool> cache\_target\_full\_ratio’ (Geoffrey Hartz)
- mon: fix quorum feature check (Greg Farnum)
- mon: fix request forwarding in mixed firefly+dumpling clusters 9#8727, Joao Eduardo Luis)
- mon: fix rule vs ruleset check in ‘osd pool set ... crush\_ruleset’ command (John Spray)
- mon: make osd ‘down’ count accurate (Sage Weil)
- mon: set ‘next commit’ in primary-affinity reply (Ilya Dryomov)
- mon: verify CRUSH features are supported by all mons (#8738, Greg Farnum)
- msgr: fix sequence negotiation during connection reset (Guang Yang)

- osd: block scrub on blocked objects (#8011, Samuel Just)
- osd: call XFS hint ioctl less often (#8241, Ilya Dryomov)
- osd: copy xattr spill out marker on clone (Haomai Wang)
- osd: fix flush of snapped objects (#8334, Samuel Just)
- osd: fix hashindex restart of merge operation (#8332, Samuel Just)
- osd: fix osdmap subscription bug causing startup hang (Greg Farnum)
- osd: fix potential null deref (#8328, Sage Weil)
- osd: fix shutdown race (#8319, Sage Weil)
- osd: handle 'none' in CRUSH results properly during peering (#8507, Samuel Just)
- osd: set no spill out marker on new objects (Greg Farnum)
- osd: skip op ordering debug checks on tiered pools (#8380, Sage Weil)
- rados: enforce 'put' alignment (Lluis Pamies-Juarez)
- rest-api: fix for 'rx' commands (Ailing Zhang)
- rgw: calc user manifest etag and fix check (#8169, #8436, Yehuda Sadeh)
- rgw: fetch attrs on multipart completion (#8452, Yehuda Sadeh, Sylvain Munaut)
- rgw: fix buffer overflow for long instance ids (#8608, Yehuda Sadeh)
- rgw: fix entity permission check on metadata put (#8428, Yehuda Sadeh)
- rgw: fix multipart retry race (#8269, Yehuda Sadeh)
- rpm: split ceph into ceph and ceph-common RPMs (Sandon Van Ness, Dan Mick)
- sysvinit: continue startin daemons after failure doing mount (#8554, Sage Weil)

For more detailed information, see [the complete changelog](#).

## v0.80.1 Firefly

This first Firefly point release fixes a few bugs, the most visible being a problem that prevents scrub from completing in some cases.

## Notable Changes

- osd: revert incomplete scrub fix (Samuel Just)

- rgw: fix stripe calculation for manifest objects (Yehuda Sadeh)
- rgw: improve handling, memory usage for abort reads (Yehuda Sadeh)
- rgw: send Swift user manifest HTTP header (Yehuda Sadeh)
- libcephfs, ceph-fuse: expose MDS session state via admin socket (Yan, Zheng)
- osd: add simple throttle for snap trimming (Sage Weil)
- monclient: fix possible hang from ill-timed monitor connection failure (Sage Weil)
- osd: fix trimming of past HitSets (Sage Weil)
- osd: fix whiteouts for non-writeback cache modes (Sage Weil)
- osd: prevent divide by zero in tiering agent (David Zafman)
- osd: prevent busy loop when tiering agent can do no work (David Zafman)

For more detailed information, see [the complete changelog](#).

## v0.80 Firefly

---

This release will form the basis for our long-term supported release Firefly, v0.80.x. The big new features are support for erasure coding and cache tiering, although a broad range of other features, fixes, and improvements have been made across the code base. Highlights include:

- *Erasure coding*: support for a broad range of erasure codes for lower storage overhead and better data durability.
- *Cache tiering*: support for creating ‘cache pools’ that store hot, recently accessed objects with automatic demotion of colder data to a base tier. Typically the cache pool is backed by faster storage devices like SSDs.
- *Primary affinity*: Ceph now has the ability to skew selection of OSDs as the “primary” copy, which allows the read workload to be cheaply skewed away from parts of the cluster without migrating any data.
- *Key/value OSD backend* (experimental): An alternative storage backend for Ceph OSD processes that puts all data in a key/value database like leveldb. This provides better performance for workloads dominated by key/value operations (like radosgw bucket indices).
- *Standalone radosgw* (experimental): The radosgw process can now run in a standalone mode without an apache (or similar) web server or fastcgi. This simplifies deployment and can improve performance.

We expect to maintain a series of stable releases based on v0.80 Firefly for as much as a year. In the meantime, development of Ceph continues with the next release, Giant, which will feature work on the CephFS distributed file system, more alternative storage backends (like RocksDB and f2fs), RDMA support, support for pyramid erasure codes, and additional functionality in the block device (RBD) like copy-on-read and multisite mirroring.

## Upgrade Sequencing

- If your existing cluster is running a version older than v0.67 Dumpling, please first upgrade to the latest Dumpling release before upgrading to v0.80 Firefly. Please refer to the [Upgrade Sequencing](#) documentation.
- We recommend adding the following to the [mon] section of your ceph.conf prior to upgrade:

```
1. mon warn on legacy crush tunables = false
```

This will prevent health warnings due to the use of legacy CRUSH placement. Although it is possible to rebalance existing data across your cluster (see the upgrade notes below), we do not normally recommend it for production environments as a large amount of data will move and there is a significant performance impact from the rebalancing.

- Upgrade daemons in the following order:

```
i. Monitors  
ii. OSDs  
iii. MDSs and/or radosgw
```

If the ceph-mds daemon is restarted first, it will wait until all OSDs have been upgraded before finishing its startup sequence. If the ceph-mon daemons are not restarted prior to the ceph-osd daemons, they will not correctly register their new capabilities with the cluster and new features may not be usable until they are restarted a second time.

- Upgrade radosgw daemons together. There is a subtle change in behavior for multipart uploads that prevents a multipart request that was initiated with a new radosgw from being completed by an old radosgw.

## Upgrading from v0.79

- OSDMap's json-formatted dump changed for keys 'full' and 'nearfull'. What was previously being outputted as 'true' or 'false' strings are now being outputted

'true' and 'false' booleans according to json syntax.

- HEALTH\_WARN on 'mon osd down out interval == 0'. Having this option set to zero on the leader acts much like having the 'noout' flag set. This warning will only be reported if the monitor getting the 'health' or 'status' request has this option set to zero.
- Monitor 'auth' commands now require the mon 'x' capability. This matches dumpling v0.67.x and earlier, but differs from emperor v0.72.x.
- A librados WATCH operation on a non-existent object now returns ENOENT; previously it did not.
- Librados interface change: As there are no partial writes, the rados\_write() and rados\_append() operations now return 0 on success like rados\_write\_full() always has. This includes the C++ interface equivalents and AIO return values for the aio variants.
- The radosgw init script (sysvinit) now requires that the 'host = ...' line in ceph.conf, if present, match the short hostname (the output of 'hostname -s'), not the fully qualified hostname or the (occasionally non-short) output of 'hostname'. Failure to adjust this when upgrading from emperor or dumpling may prevent the radosgw daemon from starting.

## Upgrading from v0.72 Emperor

---

- See notes above.
- The 'ceph -s' or 'ceph status' command's 'num\_in\_osds' field in the JSON and XML output has been changed from a string to an int.
- The recently added 'ceph mds set allow\_new\_snaps' command's syntax has changed slightly; it is now 'ceph mds set allow\_new\_snaps true'. The 'unset' command has been removed; instead, set the value to 'false'.
- The syntax for allowing snapshots is now 'mds set allow\_new\_snaps <true|false>' instead of 'mds <set unset> allow\_new\_snaps'.
- 'rbd ls' on a pool which never held rbd images now exits with code 0. It outputs nothing in plain format, or an empty list in non-plain format. This is consistent with the behavior for a pool which used to hold images, but contains none. Scripts relying on this behavior should be updated.
- The MDS requires a new OSD operation TMAP20MAP, added in this release. When upgrading, be sure to upgrade and restart the ceph-osd daemons before the ceph-mds daemon. The MDS will refuse to start if any up OSDs do not support the new feature.
- The 'ceph mds set\_max\_mds N' command is now deprecated in favor of 'ceph mds set

`max_mds N'.`

- The ‘osd pool create ...’ syntax has changed for erasure pools.
- The default CRUSH rules and layouts are now using the ‘bobtail’ tunables and defaults. Upgraded clusters using the old values will now present with a health WARN state. This can be disabled by adding ‘mon warn on legacy crush tunables = false’ to `ceph.conf` and restarting the monitors. Alternatively, you can switch to the new tunables with ‘ceph osd crush tunables firefly,’ but keep in mind that this will involve moving a *significant* portion of the data already stored in the cluster and in a large cluster may take several days to complete. We do not recommend adjusting tunables on a production cluster.
- We now default to the ‘bobtail’ CRUSH tunable values that are first supported by Ceph clients in bobtail (v0.56) and Linux kernel version v3.9. If you plan to access a newly created Ceph cluster with an older kernel client, you should use ‘ceph osd crush tunables legacy’ to switch back to the legacy behavior. Note that making that change will likely result in some data movement in the system, so adjust the setting before populating the new cluster with data.
- We now set the HASHPSPOOL flag on newly created pools (and new clusters) by default. Support for this flag first appeared in v0.64; v0.67 Dumpling is the first major release that supports it. It is first supported by the Linux kernel version v3.9. If you plan to access a newly created Ceph cluster with an older kernel or clients (e.g., librados, librbd) from a pre-dumpling Ceph release, you should add ‘osd pool default flag hashpspool = false’ to the ‘[global]’ section of your ‘`ceph.conf`’ prior to creating your monitors (e.g., after ‘`ceph-deploy new`’ but before ‘`ceph-deploy mon create ...`’).
- The configuration option ‘osd pool default crush rule’ is deprecated and replaced with ‘osd pool default crush replicated ruleset’. ‘osd pool default crush rule’ takes precedence for backward compatibility and a deprecation warning is displayed when it is used.
- As part of fix for #6796, ‘`ceph osd pool set <pool> <var> <arg>`’ now receives `<arg>` as an integer instead of a string. This affects how ‘hashpspool’ flag is set/unset: instead of ‘true’ or ‘false’, it now must be ‘0’ or ‘1’.
- The behavior of the CRUSH ‘indep’ choose mode has been changed. No ceph cluster should have been using this behavior unless someone has manually extracted a crush map, modified a CRUSH rule to replace ‘firstn’ with ‘indep’, recompiled, and reinjected the new map into the cluster. If the ‘indep’ mode is currently in use on a cluster, the rule should be modified to use ‘firstn’ instead, and the administrator should wait until any data movement completes before upgrading.
- The ‘osd dump’ command now dumps pool snaps as an array instead of an object.

## Upgrading from v0.67 Dumpling

- See notes above.
- ceph-fuse and radosgw now use the same default values for the admin socket and log file paths that the other daemons (ceph-osd, ceph-mon, etc.) do. If you run these daemons as non-root, you may need to adjust your ceph.conf to disable these options or to adjust the permissions on /var/run/ceph and /var/log/ceph.
- The MDS now disallows snapshots by default as they are not considered stable. The command ‘ceph mds set allow\_snaps’ will enable them.
- For clusters that were created before v0.44 (pre-argonaut, Spring 2012) and store radosgw data, the auto-upgrade from TMAP to OMAP objects has been disabled. Before upgrading, make sure that any buckets created on pre-argonaut releases have been modified (e.g., by PUTing and then DELETEing an object from each bucket). Any cluster created with argonaut (v0.48) or a later release or not using radosgw never relied on the automatic conversion and is not affected by this change.
- Any direct users of the ‘tmap’ portion of the librados API should be aware that the automatic tmap -> omap conversion functionality has been removed.
- Most output that used K or KB (e.g., for kilobyte) now uses a lower-case k to match the official SI convention. Any scripts that parse output and check for an upper-case K will need to be modified.
- librados::Rados::pool\_create\_async() and librados::Rados::pool\_delete\_async() don’t drop a reference to the completion object on error, caller needs to take care of that. This has never really worked correctly and we were leaking an object
- ‘ceph osd crush set <id> <weight> <loc..>’ no longer adds the osd to the specified location, as that’s a job for ‘ceph osd crush add’. It will however continue to work just the same as long as the osd already exists in the crush map.
- The OSD now enforces that class write methods cannot both mutate an object and return data. The rbd.assign\_bid method, the lone offender, has been removed. This breaks compatibility with pre-bobtail librbd clients by preventing them from creating new images.
- librados now returns on commit instead of ack for synchronous calls. This is a bit safer in the case where both OSDs and the client crash, and is probably how it should have been acting from the beginning. Users are unlikely to notice but it could result in lower performance in some circumstances. Those who care should switch to using the async interfaces, which let you specify safety semantics precisely.
- The C++ librados AioComplete::get\_version() method was incorrectly returning an int (usually 32-bits). To avoid breaking library compatibility, a get\_version64()

method is added that returns the full-width value. The old method is deprecated and will be removed in a future release. Users of the C++ librados API that make use of the get\_version() method should modify their code to avoid getting a value that is truncated from 64 to to 32 bits.

## Notable changes since v0.79

---

- ceph-fuse, libcephfs: fix several caching bugs (Yan, Zheng)
- ceph-fuse: trim inodes in response to mds memory pressure (Yan, Zheng)
- librados: fix inconsistencies in API error values (David Zafman)
- librados: fix watch operations with cache pools (Sage Weil)
- librados: new snap rollback operation (David Zafman)
- mds: fix respawn (John Spray)
- mds: misc bugs (Yan, Zheng)
- mds: misc multi-mds fixes (Yan, Zheng)
- mds: use shared\_ptr for requests (Greg Farnum)
- mon: fix peer feature checks (Sage Weil)
- mon: require 'x' mon caps for auth operations (Joao Luis)
- mon: shutdown when removed from mon cluster (Joao Luis)
- msgr: fix locking bug in authentication (Josh Durgin)
- osd: fix bug in journal replay/restart (Sage Weil)
- osd: many many many bug fixes with cache tiering (Samuel Just)
- osd: track omap and hit\_set objects in pg stats (Samuel Just)
- osd: warn if agent cannot enable due to invalid (post-split) stats (Sage Weil)
- rados bench: track metadata for multiple runs separately (Guang Yang)
- rgw: fixed subuser modify (Yehuda Sadeh)
- rpm: fix redhat-lsb dependency (Sage Weil, Alfredo Deza)

## Notable changes since v0.72 Emperor

---

- buffer: some zero-copy groundwork (Josh Durgin)

- build: misc improvements (Ken Dreyer)
- ceph-conf: stop creating bogus log files (Josh Durgin, Sage Weil)
- ceph-crush-location: new hook for setting CRUSH location of osd daemons on start)
- ceph-disk: avoid fd0 (Loic Dachary)
- ceph-disk: generalize path names, add tests (Loic Dachary)
- ceph-disk: misc improvements for puppet (Loic Dachary)
- ceph-disk: several bug fixes (Loic Dachary)
- ceph-fuse: fix race for sync reads (Sage Weil)
- ceph-fuse, libcephfs: fix several caching bugs (Yan, Zheng)
- ceph-fuse: trim inodes in response to mds memory pressure (Yan, Zheng)
- ceph-kvstore-tool: expanded command set and capabilities (Joao Eduardo Luis)
- ceph.spec: fix build dependency (Loic Dachary)
- common: bloom filter improvements (Sage Weil)
- common: check preexisting admin socket for active daemon before removing (Loic Dachary)
- common: fix aligned buffer allocation (Loic Dachary)
- common: fix authentication on big-endian architectures (Dan Mick)
- common: fix config variable substitution (Loic Dachary)
- common: portability changes to support libc++ (Noah Watkins)
- common: switch to unordered\_map from hash\_map (Noah Watkins)
- config: recursive metavariable expansion (Loic Dachary)
- crush: default to bobtail tunables (Sage Weil)
- crush: fix off-by-one error in recent refactor (Sage Weil)
- crush: many additional tests (Loic Dachary)
- crush: misc fixes, cleanups (Loic Dachary)
- crush: new rule steps to adjust retry attempts (Sage Weil)
- crush, osd: s/rep/replicated/ for less confusion (Loic Dachary)
- crush: refactor descend\_once behavior; support set\_choose\*\_tries for replicated

**rules (Sage Weil)**

- crush: usability and test improvements (Loic Dachary)
- debian: change directory ownership between ceph and ceph-common (Sage Weil)
- debian: integrate misc fixes from downstream packaging (James Page)
- doc: big update to install docs (John Wilkins)
- doc: many many install doc improvements (John Wilkins)
- doc: many many updates (John Wilkins)
- doc: misc fixes (David Moreau Simard, Kun Huang)
- erasure-code: improve buffer alignment (Loic Dachary)
- erasure-code: rewrite region-xor using vector operations (Andreas Peters)
- init: fix startup ordering/timeout problem with OSDs (Dmitry Smirnov)
- libcephfs: fix resource leak (Zheng Yan)
- librados: add C API coverage for atomic write operations (Christian Marie)
- librados: fix inconsistencies in API error values (David Zafman)
- librados: fix throttle leak (and eventual deadlock) (Josh Durgin)
- librados: fix watch operations with cache pools (Sage Weil)
- librados: new snap rollback operation (David Zafman)
- librados, osd: new TMAP20MAP operation (Yan, Zheng)
- librados: read directly into user buffer (Rutger ter Borg)
- librbd: fix use-after-free aio completion bug #5426 (Josh Durgin)
- librbd: localize/distribute parent reads (Sage Weil)
- librbd: skip zeroes/holes when copying sparse images (Josh Durgin)
- mailmap: affiliation updates (Loic Dachary)
- mailmap updates (Loic Dachary)
- many portability improvements (Noah Watkins)
- many unit test improvements (Loic Dachary)
- mds: always store backtrace in default pool (Yan, Zheng)

- mds: cope with MDS failure during creation (John Spray)
- mds: fix cap migration behavior (Yan, Zheng)
- mds: fix client session flushing (Yan, Zheng)
- mds: fix crash from client sleep/resume (Zheng Yan)
- mds: fix many many multi-mds bugs (Yan, Zheng)
- mds: fix readdir end check (Zheng Yan)
- mds: fix Resetter locking (Alexandre Oliva)
- mds: fix respawn (John Spray)
- mds: inline data support (Li Wang, Yunchuan Wen)
- mds: misc bugs (Yan, Zheng)
- mds: misc fixes for directory fragments (Zheng Yan)
- mds: misc fixes for larger directories (Zheng Yan)
- mds: misc fixes for multiple MDSs (Zheng Yan)
- mds: misc multi-mds fixes (Yan, Zheng)
- mds: remove .ceph directory (John Spray)
- mds: store directories in omap instead of tmap (Yan, Zheng)
- mds: update old-format backtraces opportunistically (Zheng Yan)
- mds: use shared\_ptr for requests (Greg Farnum)
- misc cleanups from coverity (Xing Lin)
- misc coverity fixes, cleanups (Danny Al-Gaaf)
- misc coverity fixes (Xing Lin, Li Wang, Danny Al-Gaaf)
- misc portability fixes (Noah Watkins, Alan Somers)
- misc portability fixes (Noah Watkins, Christophe Courtaut, Alan Somers, huanjun)
- misc portability work (Noah Watkins)
- mon: add erasure profiles and improve erasure pool creation (Loic Dachary)
- mon: add ‘mon getmap EPOCH’ (Joao Eduardo Luis)
- mon: allow adjustment of cephfs max file size via ‘ceph mds set max\_file\_size’ (Sage Weil)

- mon: allow debug quorum\_{enter,exit} commands via admin socket
- mon: ‘ceph osd pg-temp ...’ and primary-temp commands (Ilya Dryomov)
- mon: change mds allow\_new\_snaps syntax to be more consistent (Sage Weil)
- mon: clean up initial crush rule creation (Loic Dachary)
- mon: collect misc metadata about osd (os, kernel, etc.), new ‘osd metadata’ command (Sage Weil)
- mon: do not create erasure rules by default (Sage Weil)
- mon: do not generate spurious MDSMaps in certain cases (Sage Weil)
- mon: do not use keyring if auth = none (Loic Dachary)
- mon: fix peer feature checks (Sage Weil)
- mon: fix pg\_temp leaks (Joao Eduardo Luis)
- mon: fix pool count in ‘ceph -s’ output (Sage Weil)
- mon: handle more whitespace (newline, tab) in mon capabilities (Sage Weil)
- mon: improve (replicate or erasure) pool creation UX (Loic Dachary)
- mon: infrastructure to handle mixed-version mon cluster and cli/rest API (Greg Farnum)
- mon: MForward tests (Loic Dachary)
- mon: mkfs now idempotent (Loic Dachary)
- mon: only seed new osdmmaps to current OSDs (Sage Weil)
- mon, osd: create erasure style crush rules (Loic Dachary, Sage Weil)
- mon: ‘osd crush show-tunables’ (Sage Weil)
- mon: ‘osd dump’ dumps pool snaps as array, not object (Dan Mick)
- mon, osd: new ‘erasure’ pool type (still not fully supported)
- mon: persist quorum features to disk (Greg Farnum)
- mon: prevent extreme changes in pool pg\_num (Greg Farnum)
- mon: require ‘x’ mon caps for auth operations (Joao Luis)
- mon: shutdown when removed from mon cluster (Joao Luis)
- mon: take ‘osd pool set ...’ value as an int, not string (Joao Eduardo Luis)

- mon: track osd features in OSDMap (Joao Luis, David Zafman)
- mon: trim MDSMaps (Joao Eduardo Luis)
- mon: warn if crush has non-optimal tunables (Sage Weil)
- mount.ceph: add -n for autofs support (Steve Stock)
- msgr: fix locking bug in authentication (Josh Durgin)
- msgr: fix messenger restart race (Xihui He)
- msgr: improve connection error detection between clients and monitors (Greg Farnum, Sage Weil)
- osd: add/fix CPU feature detection for jerasure (Loic Dachary)
- osd: add HitSet tracking for read ops (Sage Weil, Greg Farnum)
- osd: avoid touching leveldb for some xattrs (Haomai Wang, Sage Weil)
- osd: backfill to multiple targets (David Zafman)
- osd: backfill to osds not in acting set (David Zafman)
- osd: cache pool support for snapshots (Sage Weil)
- osd: client IO path changes for EC (Samuel Just)
- osd: default to 3x replication
- osd: do not include backfill targets in acting set (David Zafman)
- osd: enable new hashspool layout by default (Sage Weil)
- osd: erasure plugin benchmarking tool (Loic Dachary)
- osd: fix and cleanup misc backfill issues (David Zafman)
- osd: fix bug in journal replay/restart (Sage Weil)
- osd: fix copy-get omap bug (Sage Weil)
- osd: fix linux kernel version detection (Ilya Dryomov)
- osd: fix memstore segv (Haomai Wang)
- osd: fix object\_info\_t encoding bug from emperor (Sam Just)
- osd: fix omap\_clear operation to not zap xattrs (Sam Just, Yan, Zheng)
- osd: fix several bugs with tier infrastructure
- osd: fix throttle thread (Haomai Wang)

- osd: fix XFS detection (Greg Farnum, Sushma Gurram)
- osd: generalize scrubbing infrastructure to allow EC (David Zafman)
- osd: handle more whitespace (newline, tab) in osd capabilities (Sage Weil)
- osd: ignore num\_objects\_dirty on scrub for old pools (Sage Weil)
- osd: improved scrub checks on clones (Sage Weil, Sam Just)
- osd: improve locking in fd lookup cache (Samuel Just, Greg Farnum)
- osd: include more info in pg query result (Sage Weil)
- osd, librados: fix full cluster handling (Josh Durgin)
- osd: many erasure fixes (Sam Just)
- osd: many many many bug fixes with cache tiering (Samuel Just)
- osd: move to jerasure2 library (Loic Dachary)
- osd: new 'chassis' type in default crush hierarchy (Sage Weil)
- osd: new keyvaluestore-dev backend based on leveldb (Haomai Wang)
- osd: new OSDMap encoding (Greg Farnum)
- osd: new tests for erasure pools (David Zafman)
- osd: preliminary cache pool support (no snaps) (Greg Farnum, Sage Weil)
- osd: reduce scrub lock contention (Guang Yang)
- osd: requery unfound on stray notify (#6909) (Samuel Just)
- osd: some PGBackend infrastructure (Samuel Just)
- osd: support for new 'memstore' (memory-backed) backend (Sage Weil)
- osd: track erasure compatibility (David Zafman)
- osd: track omap and hit\_set objects in pg stats (Samuel Just)
- osd: warn if agent cannot enable due to invalid (post-split) stats (Sage Weil)
- rados: add 'crush location', smart replica selection/balancing (Sage Weil)
- rados bench: track metadata for multiple runs separately (Guang Yang)
- rados: some performance optimizations (Yehuda Sadeh)
- rados tool: fix listomapvals (Josh Durgin)

- rbd: add 'rbdmap' init script for mapping rbd images on book (Adam Twardowski)
- rbd: add rbdmap support for upstart (Laurent Barbe)
- rbd: expose kernel rbd client options via 'rbd map' (Ilya Dryomov)
- rbd: fix bench-write command (Hoamai Wang)
- rbd: make 'rbd list' return empty list and success on empty pool (Josh Durgin)
- rbd: prevent deletion of images with watchers (Ilya Dryomov)
- rbd: support for 4096 mapped devices, up from ~250 (Ilya Dryomov)
- rest-api: do not fail when no OSDs yet exist (Dan Mick)
- rgw: add 'status' command to sysvinit script (David Moreau Simard)
- rgw: allow multiple frontends (Yehuda Sadeh)
- rgw: allow use of an erasure data pool (Yehuda Sadeh)
- rgw: convert bucket info to new format on demand (Yehuda Sadeh)
- rgw: fixed subuser modify (Yehuda Sadeh)
- rgw: fix error setting empty owner on ACLs (Yehuda Sadeh)
- rgw: fix fastcgi deadlock (do not return data from librados callback) (Yehuda Sadeh)
- rgw: fix many-part multipart uploads (Yehuda Sadeh)
- rgw: fix misc CORS bugs (Robin H. Johnson)
- rgw: fix object placement read op (Yehuda Sadeh)
- rgw: fix reading bucket policy (#6940)
- rgw: fix read\_user\_buckets 'max' behavior (Yehuda Sadeh)
- rgw: fix several CORS bugs (Robin H. Johnson)
- rgw: fix use-after-free when releasing completion handle (Yehuda Sadeh)
- rgw: improve swift temp URL support (Yehuda Sadeh)
- rgw: make multi-object delete idempotent (Yehuda Sadeh)
- rgw: optionally defer to bucket ACLs instead of object ACLs (Liam Monahan)
- rgw: prototype mongoose frontend (Yehuda Sadeh)
- rgw: several doc fixes (Alexandre Marangone)

- rgw: support for password (instead of admin token) for keystone authentication (Christophe Courtaut)
- rgw: switch from mongoose to civetweb (Yehuda Sadeh)
- rgw: user quotas (Yehuda Sadeh)
- rpm: fix redhat-lsb dependency (Sage Weil, Alfredo Deza)
- specfile: fix RPM build on RHEL6 (Ken Dreyer, Derek Yarnell)
- specfile: ship libdir/ceph (Key Dreyer)
- sysvinit, upstart: prevent both init systems from starting the same daemons (Josh Durgin)

## Notable changes since v0.67 Dumpling

---

- build cleanly under clang (Christophe Courtaut)
- build: Makefile refactor (Roald J. van Loon)
- build: fix [/usr]/sbin locations (Alan Somers)
- ceph-disk: fix journal preallocation
- ceph-fuse, radosgw: enable admin socket and logging by default
- ceph-fuse: fix problem with readahead vs truncate race (Yan, Zheng)
- ceph-fuse: trim deleted inodes from cache (Yan, Zheng)
- ceph-fuse: use newer fuse api (Jianpeng Ma)
- ceph-kvstore-tool: new tool for working with leveldb (copy, crc) (Joao Luis)
- ceph-post-file: new command to easily share logs or other files with ceph devs
- ceph: improve parsing of CEPH\_ARGS (Benoit Knecht)
- ceph: make -h behave when monitors are down
- ceph: parse CEPH\_ARGS env variable
- common: bloom\_filter improvements, cleanups
- common: cache crc32c values where possible
- common: correct SI is kB not KB (Dan Mick)
- common: fix looping on BSD (Alan Somers)
- common: migrate SharedPtrRegistry to use boost::shared\_ptr<> (Loic Dachary)

- common: misc portability fixes (Noah Watkins)
- crc32c: fix optimized crc32c code (it now detects arch support properly)
- crc32c: improved intel-optimized crc32c support (~8x faster on my laptop!)
- crush: fix name caching
- doc: erasure coding design notes (Loic Dachary)
- hadoop: removed old version of shim to avoid confusing users (Noah Watkins)
- librados, mon: ability to query/ping out-of-quorum monitor status (Joao Luis)
- librados: fix async aio completion wakeup
- librados: fix installed header #includes (Dan Mick)
- librados: get\_version64() method for C++ API
- librados: hello\_world example (Greg Farnum)
- librados: sync calls now return on commit (instead of ack) (Greg Farnum)
- librbd python bindings: fix parent image name limit (Josh Durgin)
- librbd, ceph-fuse: avoid some sources of ceph-fuse, rbd cache stalls
- mds: avoid leaking objects when deleting truncated files (Yan, Zheng)
- mds: fix F\_GETLK (Yan, Zheng)
- mds: fix LOOKUPSNAP bug
- mds: fix heap profiler commands (Joao Luis)
- mds: fix locking deadlock (David Disseldorp)
- mds: fix many bugs with stray (unlinked) inodes (Yan, Zheng)
- mds: fix many directory fragmentation bugs (Yan, Zheng)
- mds: fix mds rejoin with legacy parent backpointer xattrs (Alexandre Oliva)
- mds: fix rare restart/failure race during fs creation
- mds: fix standby-replay when we fall behind (Yan, Zheng)
- mds: fix stray directory purging (Yan, Zheng)
- mds: notify clients about deleted files (so they can release from their cache) (Yan, Zheng)
- mds: several bug fixes with clustered mds (Yan, Zheng)

- mon, osd: improve osdmap trimming logic (Samuel Just)
- mon, osd: initial CLI for configuring tiering
- mon: a few ‘ceph mon add’ races fixed (command is now idempotent) (Joao Luis)
- mon: allow (un)setting HASHPSPOOL flag on existing pools (Joao Luis)
- mon: allow cap strings with . to be unquoted
- mon: allow logging level of cluster log (/var/log/ceph/ceph.log) to be adjusted
- mon: avoid rewriting full osdmmaps on restart (Joao Luis)
- mon: continue to discover peer addr info during election phase
- mon: disallow CephFS snapshots until ‘ceph mds set allow\_new\_snaps’ (Greg Farnum)
- mon: do not expose uncommitted state from ‘osd crush {add,set} ...’ (Joao Luis)
- mon: fix ‘ceph osd crush reweight ...’ (Joao Luis)
- mon: fix ‘osd crush move ...’ command for buckets (Joao Luis)
- mon: fix byte counts (off by factor of 4) (Dan Mick, Joao Luis)
- mon: fix paxos corner case
- mon: kv properties for pools to support EC (Loic Dachary)
- mon: make ‘osd pool rename’ idempotent (Joao Luis)
- mon: modify ‘auth add’ semantics to make a bit more sense (Joao Luis)
- mon: new ‘osd perf’ command to dump recent performance information (Samuel Just)
- mon: new and improved ‘ceph -s’ or ‘ceph status’ command (more info, easier to read)
- mon: some auth check cleanups (Joao Luis)
- mon: track per-pool stats (Joao Luis)
- mon: warn about pools with bad pg\_num
- mon: warn when mon data stores grow very large (Joao Luis)
- monc: fix small memory leak
- new wireshark patches pulled into the tree (Kevin Jones)
- objecter, librados: redirect requests based on cache tier config
- objecter: fix possible hang when cluster is unpause (Josh Durgin)

- osd, librados: add new COPY\_FROM rados operation
- osd, librados: add new COPY\_GET rados operations (used by COPY\_FROM)
- osd: 'osd recover clone overlap limit' option to limit cloning during recovery (Samuel Just)
- osd: COPY\_GET on-wire encoding improvements (Greg Farnum)
- osd: add 'osd heartbeat min healthy ratio' configurable (was hard-coded at 33%)
- osd: add option to disable pg log debug code (which burns CPU)
- osd: allow cap strings with . to be unquoted
- osd: automatically detect proper xattr limits (David Zafman)
- osd: avoid extra copy in erasure coding reference implementation (Loic Dachary)
- osd: basic cache pool redirects (Greg Farnum)
- osd: basic whiteout, dirty flag support (not yet used)
- osd: bloom\_filter encodability, fixes, cleanups (Loic Dachary, Sage Weil)
- osd: clean up and generalize copy-from code (Greg Farnum)
- osd: cls\_hello OSD class example
- osd: erasure coding doc updates (Loic Dachary)
- osd: erasure coding plugin infrastructure, tests (Loic Dachary)
- osd: experimental support for ZFS (zfsonlinux.org) (Yan, Zheng)
- osd: fix RORDER flags
- osd: fix exponential backoff of slow request warnings (Loic Dachary)
- osd: fix handling of racing read vs write (Samuel Just)
- osd: fix version value returned by various operations (Greg Farnum)
- osd: generalized temp object infrastructure
- osd: gobject\_t infrastructure for EC (David Zafman)
- osd: improvements for compatset support and storage (David Zafman)
- osd: infrastructure to copy objects from other OSDs
- osd: instrument peering states (David Zafman)
- osd: misc copy-from improvements

- osd: opportunistic crc checking on stored data (off by default)
- osd: properly enforce RD/WR flags for rados classes
- osd: reduce blocking on backing fs (Samuel Just)
- osd: refactor recovery using PGBackend (Samuel Just)
- osd: remove old magical tmap->omap conversion
- osd: remove old pg log on upgrade (Samuel Just)
- osd: revert xattr size limit (fixes large rgw uploads)
- osd: use fdatasync(2) instead of fsync(2) to improve performance (Sam Just)
- pybind: fix blacklisting nonce (Loic Dachary)
- radosgw-agent: multi-region replication/DR
- rgw: complete in-progress requests before shutting down
- rgw: default log level is now more reasonable (Yehuda Sadeh)
- rgw: fix S3 auth with response-\* query string params (Sylvain Munaut, Yehuda Sadeh)
- rgw: fix a few minor memory leaks (Yehuda Sadeh)
- rgw: fix acl group check (Yehuda Sadeh)
- rgw: fix inefficient use of std::list::size() (Yehuda Sadeh)
- rgw: fix major CPU utilization bug with internal caching (Yehuda Sadeh, Mark Nelson)
- rgw: fix ordering of write operations (preventing data loss on crash) (Yehuda Sadeh)
- rgw: fix ordering of writes for multipart upload (Yehuda Sadeh)
- rgw: fix various CORS bugs (Yehuda Sadeh)
- rgw: fix/improve swift COPY support (Yehuda Sadeh)
- rgw: improve help output (Christophe Courtaut)
- rgw: misc fixes to support DR (Josh Durgin, Yehuda Sadeh)
- rgw: per-bucket quota (Yehuda Sadeh)
- rgw: validate S3 tokens against keystone (Roald J. van Loon)
- rgw: wildcard support for keystone roles (Christophe Courtaut)

- rpm: fix junit dependencies (Alan Grosskurth)
- sysvinit radosgw: fix status return code (Danny Al-Gaaf)
- sysvinit rbdmap: fix error ‘service rbdmap stop’ (Laurent Barbe)
- sysvinit: add condrestart command (Dan van der Ster)
- sysvinit: fix shutdown order (mons last) (Alfredo Deza)

## v0.79

---

This release is intended to serve as a release candidate for firefly, which will hopefully be v0.80. No changes are being made to the code base at this point except those that fix bugs. Please test this release if you intend to make use of the new erasure-coded pools or cache tiers in firefly.

This release fixes a range of bugs found in v0.78 and streamlines the user experience when creating erasure-coded pools. There is also a raft of fixes for the MDS (multi-mds, directory fragmentation, and large directories). The main notable new piece of functionality is a small change to allow radosgw to use an erasure-coded pool for object data.

## Upgrading

---

- Erasure pools created with v0.78 will no longer function with v0.79. You will need to delete the old pool and create a new one.
- A bug was fixed in the authentication handshake with big-endian architectures that prevent authentication between big- and little-endian machines in the same cluster. If you have a cluster that consists entirely of big-endian machines, you will need to upgrade all daemons and clients and restart.
- The ‘ceph.file.layout’ and ‘ceph.dir.layout’ extended attributes are no longer included in the listxattr(2) results to prevent problems with ‘cp -a’ and similar tools.
- Monitor ‘auth’ read-only commands now expect the user to have ‘rx’ caps. This is the same behavior that was present in dumpling, but in emperor and more recent development releases the ‘r’ cap was sufficient. The affected commands are:

1. ceph auth export
2. ceph auth get
3. ceph auth get-key
4. ceph auth print-key
5. ceph auth list

# Notable Changes

---

- ceph-conf: stop creating bogus log files (Josh Durgin, Sage Weil)
- common: fix authentication on big-endian architectures (Dan Mick)
- debian: change directory ownership between ceph and ceph-common (Sage Weil)
- init: fix startup ordering/timeout problem with OSDs (Dmitry Smirnov)
- librbd: skip zeroes/holes when copying sparse images (Josh Durgin)
- mds: cope with MDS failure during creation (John Spray)
- mds: fix crash from client sleep/resume (Zheng Yan)
- mds: misc fixes for directory fragments (Zheng Yan)
- mds: misc fixes for larger directories (Zheng Yan)
- mds: misc fixes for multiple MDSs (Zheng Yan)
- mds: remove .ceph directory (John Spray)
- misc coverity fixes, cleanups (Danny Al-Gaaf)
- mon: add erasure profiles and improve erasure pool creation (Loic Dachary)
- mon: ‘ceph osd pg-temp ...’ and primary-temp commands (Ilya Dryomov)
- mon: fix pool count in ‘ceph -s’ output (Sage Weil)
- msgr: improve connection error detection between clients and monitors (Greg Farnum, Sage Weil)
- osd: add/fix CPU feature detection for jerasure (Loic Dachary)
- osd: improved scrub checks on clones (Sage Weil, Sam Just)
- osd: many erasure fixes (Sam Just)
- osd: move to jerasure2 library (Loic Dachary)
- osd: new tests for erasure pools (David Zafman)
- osd: reduce scrub lock contention (Guang Yang)
- rgw: allow use of an erasure data pool (Yehuda Sadeh)

## v0.78

---

This development release includes two key features: erasure coding and cache tiering.

A huge amount of code was merged for this release and several additional weeks were spent stabilizing the code base, and it is now in a state where it is ready to be tested by a broader user base.

This is *not* the firefly release. Firefly will be delayed for at least another sprint so that we can get some operational experience with the new code and do some additional testing before committing to long term support.

#### Note

Please note that while it is possible to create and test erasure coded pools in this release, the pools will not be usable when you upgrade to v0.79 as the OSDMap encoding will subtly change. Please do not populate your test pools with important data that can't be reloaded.

## Upgrading

- Upgrade daemons in the following order:

- i. Monitors
- ii. OSDs
- iii. MDSs and/or radosgw

If the ceph-mds daemon is restarted first, it will wait until all OSDs have been upgraded before finishing its startup sequence. If the ceph-mon daemons are not restarted prior to the ceph-osd daemons, they will not correctly register their new capabilities with the cluster and new features may not be usable until they are restarted a second time.

- Upgrade radosgw daemons together. There is a subtle change in behavior for multipart uploads that prevents a multipart request that was initiated with a new radosgw from being completed by an old radosgw.
- CephFS recently added support for a new ‘backtrace’ attribute on file data objects that is used for lookup by inode number (i.e., NFS reexport and hard links), and will later be used by fsck repair. This replaces the existing anchor table mechanism that is used for hard link resolution. In order to completely phase that out, any inode that has an outdated backtrace attribute will get updated when the inode itself is modified. This will result in some extra workload after a legacy CephFS file system is upgraded.
- The per-op return code in librados’ ObjectWriteOperation interface is now filled in.
- The librados cmpxattr operation now handles xattrs containing null bytes as data rather than null-terminated strings.

- Compound operations in librados that create and then delete the same object are now explicitly disallowed (they fail with -EINVAL).
- The default leveldb cache size for the ceph-osd daemon has been increased from 4 MB to 128 MB. This will increase the memory footprint of that process but tends to increase performance of omap (key/value) objects (used for CephFS and the radosgw). If memory in your deployment is tight, you can preserve the old behavior by adding:

```
1. leveldb write buffer size = 0
2. leveldb cache size = 0
```

to your ceph.conf to get back the (leveldb) defaults.

## Notable Changes

---

- ceph-brag: new client and server tools (Sebastien Han, Babu Shanmugam)
- ceph-disk: use partx on RHEL or CentOS instead of partprobe (Alfredo Deza)
- ceph: fix combination of 'tell' and interactive mode (Joao Eduardo Luis)
- ceph-fuse: fix bugs with inline data and multiple MDSs (Zheng Yan)
- client: fix getcwd() to use new LOOKUPPARENT operation (Zheng Yan)
- common: fall back to json-pretty for admin socket (Loic Dachary)
- common: fix 'config dump' debug prefix (Danny Al-Gaaf)
- common: misc coverity fixes (Danny Al-Gaaf)
- common: throtller, shared\_cache performance improvements, TrackedOp (Greg Farnum, Samuel Just)
- crush: fix JSON schema for dump (John Spray)
- crush: misc cleanups, tests (Loic Dachary)
- crush: new vary\_r tunable (Sage Weil)
- crush: prevent invalid buckets of type 0 (Sage Weil)
- keyvaluestore: add perfcounters, misc bug fixes (Haomai Wang)
- keyvaluestore: portability improvements (Noah Watkins)
- libcephfs: API changes to better support NFS reexport via Ganesha (Matt Benjamin, Adam Emerson, Andrey Kuznetsov, Casey Bodley, David Zafman)
- librados: API documentation improvements (John Wilkins, Josh Durgin)

- librados: fix object enumeration bugs; allow iterator assignment (Josh Durgin)
- librados: streamline tests (Josh Durgin)
- librados: support for atomic read and omap operations for C API (Josh Durgin)
- librados: support for osd and mon command timeouts (Josh Durgin)
- librbd: pass allocation hints to OSD (Ilya Dryomov)
- logrotate: fix bug that prevented rotation for some daemons (Loic Dachary)
- mds: avoid duplicated discovers during recovery (Zheng Yan)
- mds: fix file lock owner checks (Zheng Yan)
- mds: fix LOOKUPPARENT, new LOOKUPNAME ops for reliable NFS reexport (Zheng Yan)
- mds: fix xattr handling on setxattr (Zheng Yan)
- mds: fix xattrs in getattr replies (Sage Weil)
- mds: force backtrace updates for old inodes on update (Zheng Yan)
- mds: several multi-mds and dirfrag bug fixes (Zheng Yan)
- mon: encode erasure stripe width in pool metadata (Loic Dachary)
- mon: erasure code crush rule creation (Loic Dachary)
- mon: erasure code plugin support (Loic Dachary)
- mon: fix bugs in initial post-mkfs quorum creation (Sage Weil)
- mon: fix error output to terminal during startup (Joao Eduardo Luis)
- mon: fix legacy CRUSH tunables warning (Sage Weil)
- mon: fix osd\_epochs lower bound tracking for map trimming (Sage Weil)
- mon: fix OSDMap encoding features (Sage Weil, Aaron Ten Clay)
- mon: fix ‘pg dump’ JSON output (John Spray)
- mon: include dirty stats in ‘ceph df detail’ (Sage Weil)
- mon: list quorum member names in quorum order (Sage Weil)
- mon: prevent addition of non-empty cache tier (Sage Weil)
- mon: prevent deletion of CephFS pools (John Spray)
- mon: warn when cache tier approaches ‘full’ (Sage Weil)

- osd: allocation hint, with XFS support (Ilya Dryomov)
- osd: erasure coded pool support (Samuel Just)
- osd: fix bug causing slow/stalled recovery (#7706) (Samuel Just)
- osd: fix bugs in log merging (Samuel Just)
- osd: fix/clarify end-of-object handling on read (Loic Dachary)
- osd: fix impolite mon session backoff, reconnect behavior (Greg Farnum)
- osd: fix SnapContext cache id bug (Samuel Just)
- osd: increase default leveldb cache size and write buffer (Sage Weil, Dmitry Smirnov)
- osd: limit size of 'osd bench ...' arguments (Joao Eduardo Luis)
- osdmaptool: new -test-map-pgs mode (Sage Weil, Ilya Dryomov)
- osd, mon: add primary-affinity to adjust selection of primaries (Sage Weil)
- osd: new 'status' admin socket command (Sage Weil)
- osd: simple tiering agent (Sage Weil)
- osd: store checksums for erasure coded object stripes (Samuel Just)
- osd: tests for objectstore backends (Haomai Wang)
- osd: various refactoring and bug fixes (Samuel Just, David Zafman)
- rados: add 'set-alloc-hint' command (Ilya Dryomov)
- rbd-fuse: fix enumerate\_images overflow, memory leak (Ilya Dryomov)
- rbdmap: fix upstart script (Stephan Renatus)
- rgw: avoid logging system events to usage log (Yehuda Sadeh)
- rgw: fix Swift range response (Yehuda Sadeh)
- rgw: improve scalability for manifest objects (Yehuda Sadeh)
- rgw: misc fixes for multipart objects, policies (Yehuda Sadeh)
- rgw: support non-standard MultipartUpload command (Yehuda Sadeh)

## v0.77

This is the final development release before the Firefly feature freeze. The main

items in this release include some additional refactoring work in the OSD IO path (include some locking improvements), per-user quotas for the radosgw, a switch to civetweb from mongoose for the prototype radosgw standalone mode, and a prototype leveldb-based backend for the OSD. The C librados API also got support for atomic write operations (read side transactions will appear in v0.78).

## Upgrading

---

- The ‘ceph -s’ or ‘ceph status’ command’s ‘num\_in\_osds’ field in the JSON and XML output has been changed from a string to an int.
- The recently added ‘ceph mds set allow\_new\_snaps’ command’s syntax has changed slightly; it is now ‘ceph mds set allow\_new\_snaps true’. The ‘unset’ command has been removed; instead, set the value to ‘false’.
- The syntax for allowing snapshots is now ‘mds set allow\_new\_snaps <true|false>’ instead of ‘mds <set unset> allow\_new\_snaps’.

## Notable Changes

---

- osd: client IO path changes for EC (Samuel Just)
- common: portability changes to support libc++ (Noah Watkins)
- common: switch to unordered\_map from hash\_map (Noah Watkins)
- rgw: switch from mongoose to civetweb (Yehuda Sadeh)
- osd: improve locking in fd lookup cache (Samuel Just, Greg Farnum)
- doc: many many updates (John Wilkins)
- rgw: user quotas (Yehuda Sadeh)
- mon: persist quorum features to disk (Greg Farnum)
- mon: MForward tests (Loic Dachary)
- mds: inline data support (Li Wang, Yunchuan Wen)
- rgw: fix many-part multipart uploads (Yehuda Sadeh)
- osd: new keyvaluestore-dev backend based on leveldb (Haomai Wang)
- rbd: prevent deletion of images with watchers (Ilya Dryomov)
- osd: avoid touching leveldb for some xattrs (Haomai Wang, Sage Weil)
- mailmap: affiliation updates (Loic Dachary)

- osd: new OSDMap encoding (Greg Farnum)
- osd: generalize scrubbing infrastructure to allow EC (David Zafman)
- rgw: several doc fixes (Alexandre Marangone)
- librados: add C API coverage for atomic write operations (Christian Marie)
- rgw: improve swift temp URL support (Yehuda Sadeh)
- rest-api: do not fail when no OSDs yet exist (Dan Mick)
- common: check preexisting admin socket for active daemon before removing (Loic Dachary)
- osd: handle more whitespace (newline, tab) in osd capabilities (Sage Weil)
- mon: handle more whitespace (newline, tab) in mon capabilities (Sage Weil)
- rgw: make multi-object delete idempotent (Yehuda Sadeh)
- crush: fix off-by-one error in recent refactor (Sage Weil)
- rgw: fix read\_user\_buckets 'max' behavior (Yehuda Sadeh)
- mon: change mds allow\_new\_snaps syntax to be more consistent (Sage Weil)

## v0.76

---

This release includes another batch of updates for firefly functionality. Most notably, the cache pool infrastructure now support snapshots, the OSD backfill functionality has been generalized to include multiple targets (necessary for the coming erasure pools), and there were performance improvements to the erasure code plugin on capable processors. The MDS now properly utilizes (and seamlessly migrates to) the OSD key/value interface (aka omap) for storing directory objects. There continue to be many other fixes and improvements for usability and code portability across the tree.

## Upgrading

---

- ‘rbd ls’ on a pool which never held rbd images now exits with code 0. It outputs nothing in plain format, or an empty list in non-plain format. This is consistent with the behavior for a pool which used to hold images, but contains none. Scripts relying on this behavior should be updated.
- The MDS requires a new OSD operation TMAP20MAP, added in this release. When upgrading, be sure to upgrade and restart the ceph-osd daemons before the ceph-mds daemon. The MDS will refuse to start if any up OSDs do not support the new feature.

- The ‘ceph mds set\_max\_mds N’ command is now deprecated in favor of ‘ceph mds set max\_mds N’.

## Notable Changes

---

- build: misc improvements (Ken Dreyer)
- ceph-disk: generalize path names, add tests (Loic Dachary)
- ceph-disk: misc improvements for puppet (Loic Dachary)
- ceph-disk: several bug fixes (Loic Dachary)
- ceph-fuse: fix race for sync reads (Sage Weil)
- config: recursive metavariable expansion (Loic Dachary)
- crush: usability and test improvements (Loic Dachary)
- doc: misc fixes (David Moreau Simard, Kun Huang)
- erasure-code: improve buffer alignment (Loic Dachary)
- erasure-code: rewrite region-xor using vector operations (Andreas Peters)
- librados, osd: new TMAP20MAP operation (Yan, Zheng)
- mailmap updates (Loic Dachary)
- many portability improvements (Noah Watkins)
- many unit test improvements (Loic Dachary)
- mds: always store backtrace in default pool (Yan, Zheng)
- mds: store directories in omap instead of tmap (Yan, Zheng)
- mon: allow adjustment of cephfs max file size via ‘ceph mds set max\_file\_size’ (Sage Weil)
- mon: do not create erasure rules by default (Sage Weil)
- mon: do not generate spurious MDSMaps in certain cases (Sage Weil)
- mon: do not use keyring if auth = none (Loic Dachary)
- mon: fix pg\_temp leaks (Joao Eduardo Luis)
- osd: backfill to multiple targets (David Zafman)
- osd: cache pool support for snapshots (Sage Weil)
- osd: fix and cleanup misc backfill issues (David Zafman)

- osd: fix omap\_clear operation to not zap xattrs (Sam Just, Yan, Zheng)
- osd: ignore num\_objects\_dirty on scrub for old pools (Sage Weil)
- osd: include more info in pg query result (Sage Weil)
- osd: track erasure compatibility (David Zafman)
- rbd: make 'rbd list' return empty list and success on empty pool (Josh Durgin)
- rgw: fix object placement read op (Yehuda Sadeh)
- rgw: fix several CORS bugs (Robin H. Johnson)
- specfile: fix RPM build on RHEL6 (Ken Dreyer, Derek Yarnell)
- specfile: ship libdir/ceph (Key Dreyer)

## v0.75

---

This is a big release, with lots of infrastructure going in for firefly. The big items include a prototype standalone frontend for radosgw (which does not require apache or fastcgi), tracking for read activity on the osds (to inform tiering decisions), preliminary cache pool support (no snapshots yet), and lots of bug fixes and other work across the tree to get ready for the next batch of erasure coding patches.

For comparison, here are the diff stats for the last few versions:

```
1. v0.75 291 files changed, 82713 insertions(+), 33495 deletions(-)
2. v0.74 192 files changed, 17980 insertions(+), 1062 deletions(-)
3. v0.73 148 files changed, 4464 insertions(+), 2129 deletions(-)
```

## Upgrading

---

- The 'osd pool create ...' syntax has changed for erasure pools.
- The default CRUSH rules and layouts are now using the latest and greatest tunables and defaults. Clusters using the old values will now present with a health WARN state. This can be disabled by adding 'mon warn on legacy crush tunables = false' to ceph.conf.

## Notable Changes

---

- common: bloom filter improvements (Sage Weil)
- common: fix config variable substitution (Loic Dachary)
- crush, osd: s/rep/replicated/ for less confusion (Loic Dachary)

- crush: refactor descend\_once behavior; support set\_choose\*\_tries for replicated rules (Sage Weil)
- librados: fix throttle leak (and eventual deadlock) (Josh Durgin)
- librados: read directly into user buffer (Rutger ter Borg)
- librbd: fix use-after-free aio completion bug #5426 (Josh Durgin)
- librbd: localize/distribute parent reads (Sage Weil)
- mds: fix Resetter locking (Alexandre Oliva)
- mds: fix cap migration behavior (Yan, Zheng)
- mds: fix client session flushing (Yan, Zheng)
- mds: fix many many multi-mds bugs (Yan, Zheng)
- misc portability work (Noah Watkins)
- mon, osd: create erasure style crush rules (Loic Dachary, Sage Weil)
- mon: 'osd crush show-tunables' (Sage Weil)
- mon: clean up initial crush rule creation (Loic Dachary)
- mon: improve (replicate or erasure) pool creation UX (Loic Dachary)
- mon: infrastructure to handle mixed-version mon cluster and cli/rest API (Greg Farnum)
- mon: mkfs now idempotent (Loic Dachary)
- mon: only seed new osdmmaps to current OSDs (Sage Weil)
- mon: track osd features in OSDMap (Joao Luis, David Zafman)
- mon: warn if crush has non-optimal tunables (Sage Weil)
- mount.ceph: add -n for autoofs support (Steve Stock)
- msgr: fix messenger restart race (Xihui He)
- osd, librados: fix full cluster handling (Josh Durgin)
- osd: add HitSet tracking for read ops (Sage Weil, Greg Farnum)
- osd: backfill to osds not in acting set (David Zafman)
- osd: enable new hashspool layout by default (Sage Weil)
- osd: erasure plugin benchmarking tool (Loic Dachary)

- osd: fix XFS detection (Greg Farnum, Sushma Gurram)
- osd: fix copy-get omap bug (Sage Weil)
- osd: fix linux kernel version detection (Ilya Dryomov)
- osd: fix memstore segv (Haomai Wang)
- osd: fix several bugs with tier infrastructure
- osd: fix throttle thread (Haomai Wang)
- osd: preliminary cache pool support (no snaps) (Greg Farnum, Sage Weil)
- rados tool: fix listomapvals (Josh Durgin)
- rados: add ‘crush location’, smart replica selection/balancing (Sage Weil)
- rados: some performance optimizations (Yehuda Sadeh)
- rbd: add rbdmap support for upstart (Laurent Barbe)
- rbd: expose kernel rbd client options via ‘rbd map’ (Ilya Dryomov)
- rbd: fix bench-write command (Hoamai Wang)
- rbd: support for 4096 mapped devices, up from ~250 (Ilya Dryomov)
- rgw: allow multiple frontends (Yehuda Sadeh)
- rgw: convert bucket info to new format on demand (Yehuda Sadeh)
- rgw: fix misc CORS bugs (Robin H. Johnson)
- rgw: prototype mongoose frontend (Yehuda Sadeh)

## v0.74

---

This release includes a few substantial pieces for Firefly, including a long-overdue switch to 3x replication by default and a switch to the “new” CRUSH tunables by default (supported since bobtail). There is also a fix for a long-standing radosgw bug (stalled GET) that has already been backported to emperor and dumpling.

## Upgrading

---

- We now default to the ‘bobtail’ CRUSH tunable values that are first supported by Ceph clients in bobtail (v0.56) and Linux kernel version v3.9. If you plan to access a newly created Ceph cluster with an older kernel client, you should use ‘ceph osd crush tunables legacy’ to switch back to the legacy behavior. Note that making that change will likely result in some data movement in the system, so

adjust the setting before populating the new cluster with data.

- We now set the HASHPSPOOL flag on newly created pools (and new clusters) by default. Support for this flag first appeared in v0.64; v0.67 Dumpling is the first major release that supports it. It is first supported by the Linux kernel version v3.9. If you plan to access a newly created Ceph cluster with an older kernel or clients (e.g., librados, librbd) from a pre-dumpling Ceph release, you should add ‘osd pool default flag hashpspool = false’ to the ‘[global]’ section of your ‘ceph.conf’ prior to creating your monitors (e.g., after ‘ceph-deploy new’ but before ‘ceph-deploy mon create ...’).
- The configuration option ‘osd pool default crush rule’ is deprecated and replaced with ‘osd pool default crush replicated ruleset’. ‘osd pool default crush rule’ takes precedence for backward compatibility and a deprecation warning is displayed when it is used.

## Notable Changes

---

- buffer: some zero-copy groundwork (Josh Durgin)
- ceph-disk: avoid fd0 (Loic Dachary)
- crush: default to bobtail tunables (Sage Weil)
- crush: many additional tests (Loic Dachary)
- crush: misc fixes, cleanups (Loic Dachary)
- crush: new rule steps to adjust retry attempts (Sage Weil)
- debian: integrate misc fixes from downstream packaging (James Page)
- doc: big update to install docs (John Wilkins)
- libcephfs: fix resource leak (Zheng Yan)
- misc coverity fixes (Xing Lin, Li Wang, Danny Al-Gaaf)
- misc portability fixes (Noah Watkins, Alan Somers)
- mon, osd: new ‘erasure’ pool type (still not fully supported)
- mon: add ‘mon getmap EPOCH’ (Joao Eduardo Luis)
- mon: collect misc metadata about osd (os, kernel, etc.), new ‘osd metadata’ command (Sage Weil)
- osd: default to 3x replication
- osd: do not include backfill targets in acting set (David Zafman)

- osd: new ‘chassis’ type in default crush hierarchy (Sage Weil)
- osd: requery unfound on stray notify (#6909) (Samuel Just)
- osd: some PGBackend infrastructure (Samuel Just)
- osd: support for new ‘memstore’ (memory-backed) backend (Sage Weil)
- rgw: fix fastcgi deadlock (do not return data from librados callback) (Yehuda Sadeh)
- rgw: fix reading bucket policy (#6940)
- rgw: fix use-after-free when releasing completion handle (Yehuda Sadeh)

## v0.73

---

This release, the first development release after emperor, includes many bug fixes and a few additional pieces of functionality. The first batch of larger changes will be landing in the next version, v0.74.

## Upgrading

---

- As part of fix for #6796, ‘ceph osd pool set <pool> <var> <arg>’ now receives <arg> as an integer instead of a string. This affects how ‘hashpspool’ flag is set/unset: instead of ‘true’ or ‘false’, it now must be ‘0’ or ‘1’.
- The behavior of the CRUSH ‘indep’ choose mode has been changed. No ceph cluster should have been using this behavior unless someone has manually extracted a crush map, modified a CRUSH rule to replace ‘firstn’ with ‘indep’, recompiled, and reinjected the new map into the cluster. If the ‘indep’ mode is currently in use on a cluster, the rule should be modified to use ‘firstn’ instead, and the administrator should wait until any data movement completes before upgrading.
- The ‘osd dump’ command now dumps pool snaps as an array instead of an object.
- The radosgw init script (sysvinit) now requires that the ‘host = ...’ line in ceph.conf, if present, match the short hostname (the output of ‘hostname -s’), not the fully qualified hostname or the (occasionally non-short) output of ‘hostname’. Failure to adjust this when upgrading from emperor or dumpling may prevent the radosgw daemon from starting.

## Notable Changes

---

- ceph-crush-location: new hook for setting CRUSH location of osd daemons on start
- ceph-kvstore-tool: expanded command set and capabilities (Joao Eduardo Luis)

- ceph.spec: fix build dependency (Loic Dachary)
- common: fix aligned buffer allocation (Loic Dachary)
- doc: many many install doc improvements (John Wilkins)
- mds: fix readdir end check (Zheng Yan)
- mds: update old-format backtraces opportunistically (Zheng Yan)
- misc cleanups from coverity (Xing Lin)
- misc portability fixes (Noah Watkins, Christophe Courtaut, Alan Somers, huanjun)
- mon: ‘osd dump’ dumps pool snaps as array, not object (Dan Mick)
- mon: allow debug quorum\_{enter,exit} commands via admin socket
- mon: prevent extreme changes in pool pg\_num (Greg Farnum)
- mon: take ‘osd pool set ...’ value as an int, not string (Joao Eduardo Luis)
- mon: trim MDSMaps (Joao Eduardo Luis)
- osd: fix object\_info\_t encoding bug from emperor (Sam Just)
- rbd: add ‘rbdmap’ init script for mapping rbd images on book (Adam Twardowski)
- rgw: add ‘status’ command to sysvinit script (David Moreau Simard)
- rgw: fix error setting empty owner on ACLs (Yehuda Sadeh)
- rgw: optionally defer to bucket ACLs instead of object ACLs (Liam Monahan)
- rgw: support for password (instead of admin token) for keystone authentication (Christophe Courtaut)
- sysvinit, upstart: prevent both init systems from starting the same daemons (Josh Durgin)

# v0.72.3 Emperor (pending release)

## Upgrading

- Monitor ‘auth’ read-only commands now expect the user to have ‘rx’ caps. This is the same behavior that was present in dumpling, but in emperor and more recent development releases the ‘r’ cap was sufficient. Note that this backported security fix will break mon keys that are using the following commands but do not have the ‘x’ bit in the mon capability:

```
1. ceph auth export  
2. ceph auth get  
3. ceph auth get-key  
4. ceph auth print-key  
5. ceph auth list
```

# v0.72.2 Emperor

This is the second bugfix release for the v0.72.x Emperor series. We have fixed a hang in radosgw, and fixed (again) a problem with monitor CLI compatibility with mixed version monitors. (In the future this will no longer be a problem.)

## Upgrading

- The JSON schema for the ‘osd pool set ...’ command changed slightly. Please avoid issuing this particular command via the CLI while there is a mix of v0.72.1 and v0.72.2 monitor daemons running.
- As part of fix for #6796, ‘ceph osd pool set <pool> <var> <arg>’ now receives <arg> as an integer instead of a string. This affects how ‘hashpspool’ flag is set/unset: instead of ‘true’ or ‘false’, it now must be ‘0’ or ‘1’.

## Changes

- mon: ‘osd pool set ...’ syntax change
- osd: added test for missing on-disk HEAD object
- osd: fix osd bench block size argument
- rgw: fix hang on large object GET
- rgw: fix rare use-after-free

- rgw: various DR bug fixes
- rgw: do not return error on empty owner when setting ACL
- sysvinit, upstart: prevent starting daemons using both init systems

For more detailed information, see [the complete changelog](#).

## v0.72.1 Emperor

### Important Note

When you are upgrading from Dumpling to Emperor, do not run any of the “ceph osd pool set” commands while your monitors are running separate versions. Doing so could result in inadvertently changing cluster configuration settings that exhaust compute resources in your OSDs.

### Changes

- osd: fix upgrade bug #6761
- ceph\_filestore\_tool: introduced tool to repair errors caused by #6761

This release addresses issue #6761. Upgrading to Emperor can cause reads to begin returning ENFILE (too many open files). v0.72.1 fixes that upgrade issue and adds a tool `ceph_filestore_tool` to repair osd stores affected by this bug.

To repair a cluster affected by this bug:

1. Upgrade all osd machines to v0.72.1
2. Install the `ceph-test` package on each osd machine to get `ceph_filestore_tool`
3. Stop all osd processes
4. To see all lost objects, run the following on each osd with the osd stopped and the osd data directory mounted:
 

```
ceph_filestore_tool --list-lost-objects=true --filestore-path=<path-to-osd-filestore> --journal-path=
1. <path-to-osd-journal>
```

5. To fix all lost objects, run the following on each osd with the osd stopped and the osd data directory mounted:
 

```
ceph_filestore_tool --fix-lost-objects=true --list-lost-objects=true --filestore-path=<path-to-osd-
1. filestore> --journal-path=<path-to-osd-journal>
```

6. Once lost objects have been repaired on each osd, you can restart the cluster.

Note, the ceph\_filestore\_tool performs a scan of all objects on the osd and may take some time.

## v0.72 Emperor

---

This is the fifth major release of Ceph, the fourth since adopting a 3-month development cycle. This release brings several new features, including multi-datacenter replication for the radosgw, improved usability, and lands a lot of incremental performance and internal refactoring work to support upcoming features in Firefly.

## Important Note

---

When you are upgrading from Dumpling to Emperor, do not run any of the “ceph osd pool set” commands while your monitors are running separate versions. Doing so could result in inadvertently changing cluster configuration settings that exhaust compute resources in your OSDs.

## Highlights

---

- common: improved crc32c performance
- librados: new example client and class code
- mds: many bug fixes and stability improvements
- mon: health warnings when pool pg\_num values are not reasonable
- mon: per-pool performance stats
- osd, librados: new object copy primitives
- osd: improved interaction with backend file system to reduce latency
- osd: much internal refactoring to support ongoing erasure coding and tiering support
- rgw: bucket quotas
- rgw: improved CORS support
- rgw: performance improvements
- rgw: validate S3 tokens against Keystone

Coincident with core Ceph, the Emperor release also brings:

- radosgw-agent: support for multi-datacenter replication for disaster recovery
- tgt: improved support for iSCSI via upstream tgt

Packages for both are available on ceph.com.

## Upgrade sequencing

---

There are no specific upgrade restrictions on the order or sequence of upgrading from 0.67.x Dumpling. However, you cannot run any of the “ceph osd pool set” commands while your monitors are running separate versions. Doing so could result in inadvertently changing cluster configuration settings and exhausting compute resources in your OSDs.

It is also possible to do a rolling upgrade from 0.61.x Cuttlefish, but there are ordering restrictions. (This is the same set of restrictions for Cuttlefish to Dumpling.)

1. Upgrade ceph-common on all nodes that will use the command line ‘ceph’ utility.
2. Upgrade all monitors (upgrade ceph package, restart ceph-mon daemons). This can happen one daemon or host at a time. Note that because cuttlefish and dumpling monitors can’t talk to each other, all monitors should be upgraded in relatively short succession to minimize the risk that an untimely failure will reduce availability.
3. Upgrade all osds (upgrade ceph package, restart ceph-osd daemons). This can happen one daemon or host at a time.
4. Upgrade radosgw (upgrade radosgw package, restart radosgw daemons).

## Upgrading from v0.71

---

- ceph-fuse and radosgw now use the same default values for the admin socket and log file paths that the other daemons (ceph-osd, ceph-mon, etc.) do. If you run these daemons as non-root, you may need to adjust your ceph.conf to disable these options or to adjust the permissions on /var/run/ceph and /var/log/ceph.

## Upgrading from v0.67 Dumpling

---

- ceph-fuse and radosgw now use the same default values for the admin socket and log file paths that the other daemons (ceph-osd, ceph-mon, etc.) do. If you run these daemons as non-root, you may need to adjust your ceph.conf to disable these options or to adjust the permissions on /var/run/ceph and /var/log/ceph.
- The MDS now disallows snapshots by default as they are not considered stable. The command ‘ceph mds set allow\_snaps’ will enable them.

- For clusters that were created before v0.44 (pre-argonaut, Spring 2012) and store radosgw data, the auto-upgrade from TMAP to OMAP objects has been disabled. Before upgrading, make sure that any buckets created on pre-argonaut releases have been modified (e.g., by PUTing and then DELETEing an object from each bucket). Any cluster created with argonaut (v0.48) or a later release or not using radosgw never relied on the automatic conversion and is not affected by this change.
- Any direct users of the ‘tmap’ portion of the librados API should be aware that the automatic tmap -> omap conversion functionality has been removed.
- Most output that used K or KB (e.g., for kilobyte) now uses a lower-case k to match the official SI convention. Any scripts that parse output and check for an upper-case K will need to be modified.
- librados::Rados::pool\_create\_async() and librados::Rados::pool\_delete\_async() don’t drop a reference to the completion object on error, caller needs to take care of that. This has never really worked correctly and we were leaking an object
- ‘ceph osd crush set <id> <weight> <loc..>’ no longer adds the osd to the specified location, as that’s a job for ‘ceph osd crush add’. It will however continue to work just the same as long as the osd already exists in the crush map.
- The OSD now enforces that class write methods cannot both mutate an object and return data. The rbd.assign\_bid method, the lone offender, has been removed. This breaks compatibility with pre-bobtail librbd clients by preventing them from creating new images.
- librados now returns on commit instead of ack for synchronous calls. This is a bit safer in the case where both OSDs and the client crash, and is probably how it should have been acting from the beginning. Users are unlikely to notice but it could result in lower performance in some circumstances. Those who care should switch to using the async interfaces, which let you specify safety semantics precisely.
- The C++ librados AioComplete::get\_version() method was incorrectly returning an int (usually 32-bits). To avoid breaking library compatibility, a get\_version64() method is added that returns the full-width value. The old method is deprecated and will be removed in a future release. Users of the C++ librados API that make use of the get\_version() method should modify their code to avoid getting a value that is truncated from 64 to to 32 bits.

## Notable Changes since v0.71

---

- build: fix [/usr]/sbin locations (Alan Somers)

- ceph-fuse, radosgw: enable admin socket and logging by default
- ceph: make -h behave when monitors are down
- common: cache crc32c values where possible
- common: fix looping on BSD (Alan Somers)
- librados, mon: ability to query/ping out-of-quorum monitor status (Joao Luis)
- librbd python bindings: fix parent image name limit (Josh Durgin)
- mds: avoid leaking objects when deleting truncated files (Yan, Zheng)
- mds: fix F\_GETLK (Yan, Zheng)
- mds: fix many bugs with stray (unlinked) inodes (Yan, Zheng)
- mds: fix many directory fragmentation bugs (Yan, Zheng)
- mon: allow (un)setting HASHPOOL flag on existing pools (Joao Luis)
- mon: make 'osd pool rename' idempotent (Joao Luis)
- osd: COPY\_GET on-wire encoding improvements (Greg Farnum)
- osd: bloom\_filter encodability, fixes, cleanups (Loic Dachary, Sage Weil)
- osd: fix handling of racing read vs write (Samuel Just)
- osd: reduce blocking on backing fs (Samuel Just)
- radosgw-agent: multi-region replication/DR
- rgw: fix/improve swift COPY support (Yehuda Sadeh)
- rgw: misc fixes to support DR (Josh Durgin, Yehuda Sadeh)
- rgw: per-bucket quota (Yehuda Sadeh)
- rpm: fix junit dependencies (Alan Grosskurth)

## Notable Changes since v0.67 Dumpling

---

- build cleanly under clang (Christophe Courtaut)
- build: Makefile refactor (Roald J. van Loon)
- build: fix [/usr]/sbin locations (Alan Somers)
- ceph-disk: fix journal preallocation
- ceph-fuse, radosgw: enable admin socket and logging by default

- ceph-fuse: fix problem with readahead vs truncate race (Yan, Zheng)
- ceph-fuse: trim deleted inodes from cache (Yan, Zheng)
- ceph-fuse: use newer fuse api (Jianpeng Ma)
- ceph-kvstore-tool: new tool for working with leveldb (copy, crc) (Joao Luis)
- ceph-post-file: new command to easily share logs or other files with ceph devs
- ceph: improve parsing of CEPH\_ARGS (Benoit Knecht)
- ceph: make -h behave when monitors are down
- ceph: parse CEPH\_ARGS env variable
- common: bloom\_filter improvements, cleanups
- common: cache crc32c values where possible
- common: correct SI is kB not KB (Dan Mick)
- common: fix looping on BSD (Alan Somers)
- common: migrate SharedPtrRegistry to use boost::shared\_ptr<> (Loic Dachary)
- common: misc portability fixes (Noah Watkins)
- crc32c: fix optimized crc32c code (it now detects arch support properly)
- crc32c: improved intel-optimized crc32c support (~8x faster on my laptop!)
- crush: fix name caching
- doc: erasure coding design notes (Loic Dachary)
- hadoop: removed old version of shim to avoid confusing users (Noah Watkins)
- librados, mon: ability to query/ping out-of-quorum monitor status (Joao Luis)
- librados: fix async aio completion wakeup
- librados: fix installed header #includes (Dan Mick)
- librados: get\_version64() method for C++ API
- librados: hello\_world example (Greg Farnum)
- librados: sync calls now return on commit (instead of ack) (Greg Farnum)
- librbd python bindings: fix parent image name limit (Josh Durgin)
- librbd, ceph-fuse: avoid some sources of ceph-fuse, rbd cache stalls

- mds: avoid leaking objects when deleting truncated files (Yan, Zheng)
- mds: fix F\_GETLK (Yan, Zheng)
- mds: fix LOOKUPSNAP bug
- mds: fix heap profiler commands (Joao Luis)
- mds: fix locking deadlock (David Disseldorp)
- mds: fix many bugs with stray (unlinked) inodes (Yan, Zheng)
- mds: fix many directory fragmentation bugs (Yan, Zheng)
- mds: fix mds rejoin with legacy parent backpointer xattrs (Alexandre Oliva)
- mds: fix rare restart/failure race during fs creation
- mds: fix standby-replay when we fall behind (Yan, Zheng)
- mds: fix stray directory purging (Yan, Zheng)
- mds: notify clients about deleted files (so they can release from their cache) (Yan, Zheng)
- mds: several bug fixes with clustered mds (Yan, Zheng)
- mon, osd: improve osdmap trimming logic (Samuel Just)
- mon, osd: initial CLI for configuring tiering
- mon: a few ‘ceph mon add’ races fixed (command is now idempotent) (Joao Luis)
- mon: allow (un)setting HASHPSPOOL flag on existing pools (Joao Luis)
- mon: allow cap strings with . to be unquoted
- mon: allow logging level of cluster log (/var/log/ceph/ceph.log) to be adjusted
- mon: avoid rewriting full osdmmaps on restart (Joao Luis)
- mon: continue to discover peer addr info during election phase
- mon: disallow CephFS snapshots until ‘ceph mds set allow\_new\_snaps’ (Greg Farnum)
- mon: do not expose uncommitted state from ‘osd crush {add,set} ...’ (Joao Luis)
- mon: fix ‘ceph osd crush reweight ...’ (Joao Luis)
- mon: fix ‘osd crush move ...’ command for buckets (Joao Luis)
- mon: fix byte counts (off by factor of 4) (Dan Mick, Joao Luis)
- mon: fix paxos corner case

- mon: kv properties for pools to support EC (Loic Dachary)
- mon: make 'osd pool rename' idempotent (Joao Luis)
- mon: modify 'auth add' semantics to make a bit more sense (Joao Luis)
- mon: new 'osd perf' command to dump recent performance information (Samuel Just)
- mon: new and improved 'ceph -s' or 'ceph status' command (more info, easier to read)
- mon: some auth check cleanups (Joao Luis)
- mon: track per-pool stats (Joao Luis)
- mon: warn about pools with bad pg\_num
- mon: warn when mon data stores grow very large (Joao Luis)
- monc: fix small memory leak
- new wireshark patches pulled into the tree (Kevin Jones)
- objecter, librados: redirect requests based on cache tier config
- objecter: fix possible hang when cluster is unpause (Josh Durgin)
- osd, librados: add new COPY\_FROM rados operation
- osd, librados: add new COPY\_GET rados operations (used by COPY\_FROM)
- osd: 'osd recover clone overlap limit' option to limit cloning during recovery (Samuel Just)
- osd: COPY\_GET on-wire encoding improvements (Greg Farnum)
- osd: add 'osd heartbeat min healthy ratio' configurable (was hard-coded at 33%)
- osd: add option to disable pg log debug code (which burns CPU)
- osd: allow cap strings with . to be unquoted
- osd: automatically detect proper xattr limits (David Zafman)
- osd: avoid extra copy in erasure coding reference implementation (Loic Dachary)
- osd: basic cache pool redirects (Greg Farnum)
- osd: basic whiteout, dirty flag support (not yet used)
- osd: bloom\_filter encodability, fixes, cleanups (Loic Dachary, Sage Weil)
- osd: clean up and generalize copy-from code (Greg Farnum)

- osd: cls\_hello OSD class example
- osd: erasure coding doc updates (Loic Dachary)
- osd: erasure coding plugin infrastructure, tests (Loic Dachary)
- osd: experimental support for ZFS (zfsonlinux.org) (Yan, Zheng)
- osd: fix RWORDER flags
- osd: fix exponential backoff of slow request warnings (Loic Dachary)
- osd: fix handling of racing read vs write (Samuel Just)
- osd: fix version value returned by various operations (Greg Farnum)
- osd: generalized temp object infrastructure
- osd: gobject\_t infrastructure for EC (David Zafman)
- osd: improvements for compatset support and storage (David Zafman)
- osd: infrastructure to copy objects from other OSDs
- osd: instrument peering states (David Zafman)
- osd: misc copy-from improvements
- osd: opportunistic crc checking on stored data (off by default)
- osd: properly enforce RD/WR flags for rados classes
- osd: reduce blocking on backing fs (Samuel Just)
- osd: refactor recovery using PGBackend (Samuel Just)
- osd: remove old magical tmap->omap conversion
- osd: remove old pg log on upgrade (Samuel Just)
- osd: revert xattr size limit (fixes large rgw uploads)
- osd: use fdatasync(2) instead of fsync(2) to improve performance (Sam Just)
- pybind: fix blacklisting nonce (Loic Dachary)
- radosgw-agent: multi-region replication/DR
- rgw: complete in-progress requests before shutting down
- rgw: default log level is now more reasonable (Yehuda Sadeh)
- rgw: fix S3 auth with response-\* query string params (Sylvain Munaut, Yehuda Sadeh)

- rgw: fix a few minor memory leaks (Yehuda Sadeh)
- rgw: fix acl group check (Yehuda Sadeh)
- rgw: fix inefficient use of std::list::size() (Yehuda Sadeh)
- rgw: fix major CPU utilization bug with internal caching (Yehuda Sadeh, Mark Nelson)
- rgw: fix ordering of write operations (preventing data loss on crash) (Yehuda Sadeh)
- rgw: fix ordering of writes for multipart upload (Yehuda Sadeh)
- rgw: fix various CORS bugs (Yehuda Sadeh)
- rgw: fix/improve swift COPY support (Yehuda Sadeh)
- rgw: improve help output (Christophe Courtaut)
- rgw: misc fixes to support DR (Josh Durgin, Yehuda Sadeh)
- rgw: per-bucket quota (Yehuda Sadeh)
- rgw: validate S3 tokens against keystone (Roald J. van Loon)
- rgw: wildcard support for keystone roles (Christophe Courtaut)
- rpm: fix junit dependencies (Alan Grosskurth)
- sysvinit radosgw: fix status return code (Danny Al-Gaaf)
- sysvinit rbdmap: fix error ‘service rbdmap stop’ (Laurent Barbe)
- sysvinit: add condrestart command (Dan van der Ster)
- sysvinit: fix shutdown order (mons last) (Alfredo Deza)

## v0.71

---

This development release includes a significant amount of new code and refactoring, as well as a lot of preliminary functionality that will be needed for erasure coding and tiering support. There are also several significant patch sets improving this with the MDS.

## Upgrading

---

- The MDS now disallows snapshots by default as they are not considered stable. The command ‘ceph mds set allow\_snaps’ will enable them.

- For clusters that were created before v0.44 (pre-argonaut, Spring 2012) and store radosgw data, the auto-upgrade from TMAP to OMAP objects has been disabled. Before upgrading, make sure that any buckets created on pre-argonaut releases have been modified (e.g., by PUTing and then DELETEing an object from each bucket). Any cluster created with argonaut (v0.48) or a later release or not using radosgw never relied on the automatic conversion and is not affected by this change.
- Any direct users of the ‘tmap’ portion of the librados API should be aware that the automatic tmap -> omap conversion functionality has been removed.
- Most output that used K or KB (e.g., for kilobyte) now uses a lower-case k to match the official SI convention. Any scripts that parse output and check for an upper-case K will need to be modified.

## Notable Changes

---

- build: Makefile refactor (Roald J. van Loon)
- ceph-disk: fix journal preallocation
- ceph-fuse: trim deleted inodes from cache (Yan, Zheng)
- ceph-fuse: use newer fuse api (Jianpeng Ma)
- ceph-kvstore-tool: new tool for working with leveldb (copy, crc) (Joao Luis)
- common: bloom\_filter improvements, cleanups
- common: correct SI is kB not KB (Dan Mick)
- common: misc portability fixes (Noah Watkins)
- hadoop: removed old version of shim to avoid confusing users (Noah Watkins)
- librados: fix installed header #includes (Dan Mick)
- librbd, ceph-fuse: avoid some sources of ceph-fuse, rbd cache stalls
- mds: fix LOOKUPSNAP bug
- mds: fix standby-replay when we fall behind (Yan, Zheng)
- mds: fix stray directory purging (Yan, Zheng)
- mon: disallow CephFS snapshots until ‘ceph mds set allow\_new\_snaps’ (Greg Farnum)
- mon, osd: improve osdmap trimming logic (Samuel Just)
- mon: kv properties for pools to support EC (Loic Dachary)

- mon: some auth check cleanups (Joao Luis)
- mon: track per-pool stats (Joao Luis)
- mon: warn about pools with bad pg\_num
- osd: automatically detect proper xattr limits (David Zafman)
- osd: avoid extra copy in erasure coding reference implementation (Loic Dachary)
- osd: basic cache pool redirects (Greg Farnum)
- osd: basic whiteout, dirty flag support (not yet used)
- osd: clean up and generalize copy-from code (Greg Farnum)
- osd: erasure coding doc updates (Loic Dachary)
- osd: erasure coding plugin infrastructure, tests (Loic Dachary)
- osd: fix RWORDER flags
- osd: fix exponential backoff of slow request warnings (Loic Dachary)
- osd: generalized temp object infrastructure
- osd: ghobject\_t infrastructure for EC (David Zafman)
- osd: improvements for compatset support and storage (David Zafman)
- osd: misc copy-from improvements
- osd: opportunistic crc checking on stored data (off by default)
- osd: refactor recovery using PGBackend (Samuel Just)
- osd: remove old magical tmap->omap conversion
- pybind: fix blacklisting nonce (Loic Dachary)
- rgw: default log level is now more reasonable (Yehuda Sadeh)
- rgw: fix acl group check (Yehuda Sadeh)
- sysvinit: fix shutdown order (mons last) (Alfredo Deza)

## v0.70

---

## Upgrading

---

- librados::Rados::pool\_create\_async() and librados::Rados::pool\_delete\_async()

don't drop a reference to the completion object on error, caller needs to take care of that. This has never really worked correctly and we were leaking an object

- 'ceph osd crush set <id> <weight> <loc..>' no longer adds the osd to the specified location, as that's a job for 'ceph osd crush add'. It will however continue to work just the same as long as the osd already exists in the crush map.

## Notable Changes

---

- mon: a few 'ceph mon add' races fixed (command is now idempotent) (Joao Luis)
- crush: fix name caching
- rgw: fix a few minor memory leaks (Yehuda Sadeh)
- ceph: improve parsing of CEPH\_ARGS (Benoit Knecht)
- mon: avoid rewriting full osdmaps on restart (Joao Luis)
- crc32c: fix optimized crc32c code (it now detects arch support properly)
- mon: fix 'ceph osd crush reweight ...' (Joao Luis)
- osd: revert xattr size limit (fixes large rgw uploads)
- mds: fix heap profiler commands (Joao Luis)
- rgw: fix inefficient use of std::list::size() (Yehuda Sadeh)

## v0.69

---

## Upgrading

---

- The sysvinit /etc/init.d/ceph script will, by default, update the CRUSH location of an OSD when it starts. Previously, if the monitors were not available, this command would hang indefinitely. Now, that step will time out after 10 seconds and the ceph-osd daemon will not be started.
- Users of the librados C++ API should replace users of get\_version() with get\_version64() as the old method only returns a 32-bit value for a 64-bit field. The existing 32-bit get\_version() method is now deprecated.
- The OSDs are now more picky that request payload match their declared size. A write operation across N bytes that includes M bytes of data will now be rejected. No known clients do this, but because the server-side behavior has changed it is possible that an application misusing the interface may now get

errors.

- The OSD now enforces that class write methods cannot both mutate an object and return data. The rbd.assign\_bid method, the lone offender, has been removed. This breaks compatibility with pre-bobtail librbd clients by preventing them from creating new images.
- librados now returns on commit instead of ack for synchronous calls. This is a bit safer in the case where both OSDs and the client crash, and is probably how it should have been acting from the beginning. Users are unlikely to notice but it could result in lower performance in some circumstances. Those who care should switch to using the async interfaces, which let you specify safety semantics precisely.
- The C++ librados AioComplete::get\_version() method was incorrectly returning an int (usually 32-bits). To avoid breaking library compatibility, a get\_version64() method is added that returns the full-width value. The old method is deprecated and will be removed in a future release. Users of the C++ librados API that make use of the get\_version() method should modify their code to avoid getting a value that is truncated from 64 to 32 bits.

## Notable Changes

---

- build cleanly under clang (Christophe Courtaut)
- common: migrate SharedPtrRegistry to use boost::shared\_ptr<> (Loic Dachary)
- doc: erasure coding design notes (Loic Dachary)
- improved intel-optimized crc32c support (~8x faster on my laptop!)
- librados: get\_version64() method for C++ API
- mds: fix locking deadlock (David Disseldorp)
- mon, osd: initial CLI for configuring tiering
- mon: allow cap strings with . to be unquoted
- mon: continue to discover peer addr info during election phase
- mon: fix ‘osd crush move ...’ command for buckets (Joao Luis)
- mon: warn when mon data stores grow very large (Joao Luis)
- objecter, librados: redirect requests based on cache tier config
- osd, librados: add new COPY\_FROM rados operation
- osd, librados: add new COPY\_GET rados operations (used by COPY\_FROM)

- osd: add ‘osd heartbeat min healthy ratio’ configurable (was hard-coded at 33%)
- osd: add option to disable pg log debug code (which burns CPU)
- osd: allow cap strings with . to be unquoted
- osd: fix version value returned by various operations (Greg Farnum)
- osd: infrastructure to copy objects from other OSDs
- osd: use fdatasync(2) instead of fsync(2) to improve performance (Sam Just)
- rgw: fix major CPU utilization bug with internal caching (Yehuda Sadeh, Mark Nelson)
- rgw: fix ordering of write operations (preventing data loss on crash) (Yehuda Sadeh)
- rgw: fix ordering of writes for multipart upload (Yehuda Sadeh)
- rgw: fix various CORS bugs (Yehuda Sadeh)
- rgw: improve help output (Christophe Courtaut)
- rgw: validate S3 tokens against keystone (Roald J. van Loon)
- rgw: wildcard support for keystone roles (Christophe Courtaut)
- sysvinit radosgw: fix status return code (Danny Al-Gaaf)
- sysvinit rbdmap: fix error ‘service rbdmap stop’ (Laurent Barbe)

## v0.68

---

## Upgrading

---

- ‘ceph osd crush set <id> <weight> <loc..>’ no longer adds the osd to the specified location, as that’s a job for ‘ceph osd crush add’. It will however continue to work just the same as long as the osd already exists in the crush map.
- The OSD now enforces that class write methods cannot both mutate an object and return data. The rbd.assign\_bid method, the lone offender, has been removed. This breaks compatibility with pre-boottail librbd clients by preventing them from creating new images.
- librados now returns on commit instead of ack for synchronous calls. This is a bit safer in the case where both OSDs and the client crash, and is probably how it should have been acting from the beginning. Users are unlikely to notice but

it could result in lower performance in some circumstances. Those who care should switch to using the async interfaces, which let you specify safety semantics precisely.

- The C++ librados AioComplete::get\_version() method was incorrectly returning an int (usually 32-bits). To avoid breaking library compatibility, a get\_version64() method is added that returns the full-width value. The old method is deprecated and will be removed in a future release. Users of the C++ librados API that make use of the get\_version() method should modify their code to avoid getting a value that is truncated from 64 to to 32 bits.

## Notable Changes

---

- ceph-fuse: fix problem with readahead vs truncate race (Yan, Zheng)
- ceph-post-file: new command to easily share logs or other files with ceph devs
- ceph: parse CEPH\_ARGS env variable
- librados: fix aio completion wakeup
- librados: hello\_world example (Greg Farnum)
- librados: sync calls now return on commit (instead of ack) (Greg Farnum)
- mds: fix mds rejoin with legacy parent backpointer xattrs (Alexandre Oliva)
- mds: fix rare restart/failure race during fs creation
- mds: notify clients about deleted files (so they can release from their cache) (Yan, Zheng)
- mds: several bug fixes with clustered mds (Yan, Zheng)
- mon: allow logging level of cluster log (/var/log/ceph/ceph.log) to be adjusted
- mon: do not expose uncommitted state from ‘osd crush {add,set} ...’ (Joao Luis)
- mon: fix byte counts (off by factor of 4) (Dan Mick, Joao Luis)
- mon: fix paxos corner case
- mon: modify ‘auth add’ semantics to make a bit more sense (Joao Luis)
- mon: new ‘osd perf’ command to dump recent performance information (Samuel Just)
- mon: new and improved ‘ceph -s’ or ‘ceph status’ command (more info, easier to read)
- monc: fix small memory leak

- new wireshark patches pulled into the tree (Kevin Jones)
- objecter: fix possible hang when cluster is unpause (Josh Durgin)
- osd: 'osd recover clone overlap limit' option to limit cloning during recovery (Samuel Just)
- osd: cls\_hello OSD class example
- osd: experimental support for ZFS (zfsonlinux.org) (Yan, Zheng)
- osd: instrument peering states (David Zafman)
- osd: properly enforce RD/WR flags for rados classes
- osd: remove old pg log on upgrade (Samuel Just)
- rgw: complete in-progress requests before shutting down
- rgw: fix S3 auth with response-\* query string params (Sylvain Munaut, Yehuda Sadeh)
- sysvinit: add condrestart command (Dan van der Ster)

# v0.67.12 “Dumpling” (draft)

This stable update for Dumpling fixes a few longstanding issues with backfill in the OSD that can lead to stalled IOs. There is also a fix for memory utilization for reads in librbd when caching is enabled, and then several other small fixes across the rest of the system.

Dumpling users who have encountered IO stalls during backfill and who do not expect to upgrade to Firefly soon should upgrade. Everyone else should upgrade to Firefly already. This is likely to be the last stable release for the 0.67.x Dumpling series.

## Notable Changes

- buffer: fix buffer rebuild alignment corner case (#6614 #6003 Loic Dachary, Samuel Just)
- ceph-disk: reprobe partitions after zap (#9665 #9721 Loic Dachary)
- ceph-disk: use partx instead of partprobe when appropriate (Loic Dachary)
- common: add \$cctid meta variable (#6228 Adam Crume)
- crush: fix get\_full\_location\_ordered (Sage Weil)
- crush: pick ruleset id that matches rule\_id (#9675 Xiaoxi Chen)
- libcephfs: fix tid wrap bug (#9869 Greg Farnum)
- libcephfs: get osd location on -1 should return EINVAL (Sage Weil)
- librados: fix race condition with C API and op timeouts (#9582 Sage Weil)
- librbd: constrain max number of in-flight read requests (#9854 Jason Dillaman)
- librbd: enforce cache size on read requests (Jason Dillaman)
- librbd: fix invalid close in image open failure path (#10030 Jason Dillaman)
- librbd: fix read hang on sparse files (Jason Dillaman)
- librbd: gracefully handle deleted/renamed pools (#10270 #10122 Jason Dillaman)
- librbd: protect list\_children from invalid child pool ioctxs (#10123 Jason Dillaman)
- mds: fix ctime updates from clients without dirty caps (#9514 Greg Farnum)
- mds: fix rare NULL dereference in cap update path (Greg Farnum)

- mds: fix assertion caused by system clock backwards (#11053 Yan, Zheng)
- mds: store backtrace on straydir (Yan, Zheng)
- osd: fix journal committed\_thru update after replay (#6756 Samuel Just)
- osd: fix memory leak, busy loop on snap trim (#9113 Samuel Just)
- osd: fix misc peering, recovery bugs (#10168 Samuel Just)
- osd: fix purged\_snap field on backfill start (#9487 Sage Weil, Samuel Just)
- osd: handle no-op write with snapshot corner case (#10262 Sage Weil, Loic Dachary)
- osd: respect RWORDERED rados flag (Sage Weil)
- osd: several backfill fixes and refactors (Samuel Just, David Zafman)
- rgw: send http status reason explicitly in fastcgi (Yehuda Sadeh)

## v0.67.11 “Dumpling”

---

This stable update for Dumpling fixes several important bugs that affect a small set of users.

We recommend that all Dumpling users upgrade at their convenience. If none of these issues are affecting your deployment there is no urgency.

## Notable Changes

---

- common: fix sending dup cluster log items (#9080 Sage Weil)
- doc: several doc updates (Alfredo Deza)
- libcephfs-java: fix build against older JNI headers (Greg Farnum)
- librados: fix crash in op timeout path (#9362 Matthias Kiefer, Sage Weil)
- librbd: fix crash using clone of flattened image (#8845 Josh Durgin)
- librbd: fix error path cleanup when failing to open image (#8912 Josh Durgin)
- mon: fix crash when adjusting pg\_num before any OSDs are added (#9052 Sage Weil)
- mon: reduce log noise from paxos (Aanchal Agrawal, Sage Weil)
- osd: allow scrub and snap trim thread pool IO priority to be adjusted (Sage Weil)
- osd: fix mount/remount sync race (#9144 Sage Weil)

# v0.67.10 “Dumpling”

This stable update release for Dumpling includes primarily fixes for RGW, including several issues with bucket listings and a potential data corruption problem when multiple multi-part uploads race. There is also some throttling capability added in the OSD for scrub that can mitigate the performance impact on production clusters.

We recommend that all Dumpling users upgrade at their convenience.

## Notable Changes

- ceph-disk: partprobe before settle, fixing dm-crypt (#6966, Eric Eastman)
- librbd: add invalidate cache interface (Josh Durgin)
- librbd: close image if remove\_child fails (Ilya Dryomov)
- librbd: fix potential null pointer dereference (Danny Al-Gaaf)
- librbd: improve writeback checks, performance (Haomai Wang)
- librbd: skip zeroes when copying image (#6257, Josh Durgin)
- mon: fix rule(set) check on ‘ceph pool set ... crush\_ruleset ...’ (#8599, John Spray)
- mon: shut down if mon is removed from cluster (#6789, Joao Eduardo Luis)
- osd: fix filestore perf reports to mon (Sage Weil)
- osd: force any new or updated xattr into leveldb if E2BIG from XFS (#7779, Sage Weil)
- osd: lock snapdir object during write to fix race with backfill (Samuel Just)
- osd: option sleep during scrub (Sage Weil)
- osd: set io priority on scrub and snap trim threads (Sage Weil)
- osd: ‘status’ admin socket command (Sage Weil)
- rbd: tolerate missing NULL terminator on block\_name\_prefix (#7577, Dan Mick)
- rgw: calculate user manifest (#8169, Yehuda Sadeh)
- rgw: fix abort on chunk read error, avoid using extra memory (#8289, Yehuda Sadeh)
- rgw: fix buffer overflow on bucket instance id (#8608, Yehuda Sadeh)
- rgw: fix crash in swift CORS preflight request (#8586, Yehuda Sadeh)

- rgw: fix implicit removal of old objects on object creation (#8972, Patrycja Szablowska, Yehuda Sadeh)
- rgw: fix MaxKeys in bucket listing (Yehuda Sadeh)
- rgw: fix race with multiple updates to a single multipart object (#8269, Yehuda Sadeh)
- rgw: improve bucket listing with delimiter (Yehuda Sadeh)
- rgw: include NextMarker in bucket listing (#8858, Yehuda Sadeh)
- rgw: return error early on non-existent bucket (#7064, Yehuda Sadeh)
- rgw: set truncation flag correctly in bucket listing (Yehuda Sadeh)
- sysvinit: continue starting daemons after pre-mount error (#8554, Sage Weil)

For more detailed information, see [the complete changelog](#).

## v0.67.9 “Dumpling”

---

This Dumpling point release fixes several minor bugs. The most prevalent in the field is one that occasionally prevents OSDs from starting on recently created clusters.

We recommend that all Dumpling users upgrade at their convenience.

## Notable Changes

---

- ceph-fuse, libcephfs: client admin socket command to kick and inspect MDS sessions (#8021, Zheng Yan)
- monclient: fix failure detection during mon handshake (#8278, Sage Weil)
- mon: set tid on no-op PGStatsAck messages (#8280, Sage Weil)
- msgr: fix a rare bug with connection negotiation between OSDs (Guang Yang)
- osd: allow snap trim throttling with simple delay (#6278, Sage Weil)
- osd: check for splitting when processing recover/backfill reservations (#6565, Samuel Just)
- osd: fix backfill position tracking (#8162, Samuel Just)
- osd: fix bug in backfill stats (Samuel Just)
- osd: fix bug preventing OSD startup for infant clusters (#8162, Greg Farnum)
- osd: fix rare PG resurrection race causing an incomplete PG (#7740, Samuel Just)

- osd: only complete replicas count toward min\_size (#7805, Samuel Just)
- rgw: allow setting ACLs with empty owner (#6892, Yehuda Sadeh)
- rgw: send user manifest header field (#8170, Yehuda Sadeh)

For more detailed information, see [the complete changelog](#).

## v0.67.8 “Dumpling”

---

This Dumpling point release fixes several non-critical issues since v0.67.7. The most notable bug fixes are an auth fix in librbd (observed as an occasional crash from KVM), an improvement in the network failure detection with the monitor, and several hard to hit OSD crashes or hangs.

We recommend that all users upgrade at their convenience.

## Upgrading

---

- The ‘rbd ls’ function now returns success and returns an empty when a pool does not store any rbd images. Previously it would return an ENOENT error.
- Ceph will now issue a health warning if the ‘mon osd down out interval’ config option is set to zero. This warning can be disabled by adding ‘mon warn on osd down out interval zero = false’ to ceph.conf.

## Notable Changes

---

- all: improve keepalive detection of failed monitor connections (#7888, Sage Weil)
- ceph-fuse, libcephfs: pin inodes during readahead, fixing rare crash (#7867, Sage Weil)
- librbd: make cache writeback a bit less aggressive (Sage Weil)
- librbd: make symlink for qemu to detect librbd in RPM (#7293, Josh Durgin)
- mon: allow ‘hashpspool’ pool flag to be set and unset (Loic Dachary)
- mon: commit paxos state only after entire quorum acks, fixing rare race where prior round state is readable (#7736, Sage Weil)
- mon: make elections and timeouts a bit more robust (#7212, Sage Weil)
- mon: prevent extreme pool split operations (Greg Farnum)
- mon: wait for quorum for get\_version requests to close rare pool creation race (#7997, Sage Weil)

- mon: warn on ‘mon osd down out interval = 0’ (#7784, Joao Luis)
- msgr: fix byte-order for auth challenge, fixing auth errors on big-endian clients (#7977, Dan Mick)
- msgr: fix occasional crash in authentication code (usually triggered by librbd) (#6840, Josh Durgin)
- msgr: fix rebind() race (#6992, Xihui He)
- osd: avoid timeouts during slow PG deletion (#6528, Samuel Just)
- osd: fix bug in pool listing during recovery (#6633, Samuel Just)
- osd: fix queue limits, fixing recovery stalls (#7706, Samuel Just)
- osd: fix rare peering crashes (#6722, #6910, Samuel Just)
- osd: fix rare recovery hang (#6681, Samuel Just)
- osd: improve error handling on journal errors (#7738, Sage Weil)
- osd: reduce load on the monitor from OSDMap subscriptions (Greg Farnum)
- osd: rery GetLog on peer osd startup, fixing some rare peering stalls (#6909, Samuel Just)
- osd: reset journal state on remount to fix occasional crash on OSD startup (#8019, Sage Weil)
- osd: share maps with peers more aggressively (Greg Farnum)
- rbd: make it harder to delete an rbd image that is currently in use (#7076, Ilya Drymov)
- rgw: deny writes to secondary zone by non-system users (#6678, Yehuda Sadeh)
- rgw: do'nt log system requests in usage log (#6889, Yehuda Sadeh)
- rgw: fix bucket recreation (#6951, Yehuda Sadeh)
- rgw: fix Swift range response (#7099, Julien Calvet, Yehuda Sadeh)
- rgw: fix URL escaping (#8202, Yehuda Sadeh)
- rgw: fix whitespace trimming in http headers (#7543, Yehuda Sadeh)
- rgw: make multi-object deletion idempotent (#7346, Yehuda Sadeh)

For more detailed information, see [the complete changelog](#).

## v0.67.7 “Dumpling”

This Dumpling point release fixes a few critical issues in v0.67.6.

All v0.67.6 users are urgently encouraged to upgrade. We also recommend that all v0.67.5 (or older) users upgrade.

## Upgrading

---

- Once you have upgraded a radosgw instance or OSD to v0.67.7, you should not downgrade to a previous version.

## Notable Changes

---

- ceph-disk: additional unit tests
- librbd: revert caching behavior change in v0.67.6
- osd: fix problem reading xattrs due to incomplete backport in v0.67.6
- radosgw-admin: fix reading object policy

For more detailed information, see [the complete changelog](#).

## v0.67.6 “Dumpling”

---

This Dumpling point release contains a number of important fixes for the OSD, monitor, and radosgw. Most significantly, a change that forces large object attributes to spill over into leveldb has been backported that can prevent objects and the cluster from being damaged by large attributes (which can be induced via the radosgw). There is also a set of fixes that improves data safety and RADOS semantics when the cluster becomes full and then non-full.

We recommend that all 0.67.x Dumpling users skip this release and upgrade to v0.67.7.

## Upgrading

---

- The OSD has long contained a feature that allows large xattrs to spill over into the leveldb backing store in situations where not all local file systems are able to store them reliably. This option is now enabled unconditionally in order to avoid rare cases where storing large xattrs renders the object unreadable. This is known to be triggered by very large multipart objects, but could be caused by other workloads as well. Although there is some small risk that performance for certain workloads will degrade, it is more important that data be retrievable. Note that newer versions of Ceph (e.g., firefly) do some additional work to avoid the potential performance regression in this case, but that is currently considered too complex for backport to the Dumpling stable series.

- It is very dangerous to downgrade from v0.67.6 to a prior version of Dumpling. If the old version does not have ‘filestore xattr use omap = true’ it may not be able to read all xattrs for an object and can cause undefined behavior.

## Notable changes

---

- ceph-disk: misc bug fixes, particularly on RHEL (Loic Dachary, Alfredo Deza, various)
- ceph-fuse, libcephfs: fix crash from read over certain sparseness patterns (Sage Weil)
- ceph-fuse, libcephfs: fix integer overflow for sync reads racing with appends (Sage Weil)
- ceph.spec: fix udev rule when building RPM under RHEL (Derek Yarnell)
- common: fix crash from bad format from admin socket (Loic Dachary)
- librados: add optional timeouts (Josh Durgin)
- librados: do not leak budget when resending localized or redirected ops (Josh Durgin)
- librados, osd: fix and improve full cluster handling (Josh Durgin)
- librbd: fix use-after-free when updating perfcounters during image close (Josh Durgin)
- librbd: remove limit on objects in cache (Josh Durgin)
- mon: avoid on-disk full OSDMap corruption from pg\_temp removal (Sage Weil)
- mon: avoid stray pg\_temp entries from pool deletion race (Joao Eduardo Luis)
- mon: do not generate spurious MDSMaps from laggy daemons (Joao Eduardo Luis)
- mon: fix error code from ‘osd rm|down|out|in ...’ commands (Loic Dachary)
- mon: include all health items in summary output (John Spray)
- osd: fix occasional race/crash during startup (Sage Weil)
- osd: ignore stray OSDMap messages during init (Sage Weil)
- osd: unconditionally let xattrs overflow into leveldb (David Zafman)
- rados: fix a few error checks for the CLI (Josh Durgin)
- rgw: convert legacy bucket info objects on demand (Yehuda Sadeh)
- rgw: fix bug causing system users to lose privileges (Yehuda Sadeh)

- rgw: fix CORS bugs related to headers and case sensitivity (Robin H. Johnson)
- rgw: fix multipart object listing (Yehuda Sadeh)
- rgw: fix racing object creations (Yehuda Sadeh)
- rgw: fix racing object put and delete (Yehuda Sadeh)
- rgw: fix S3 auth when using response-\* query string params (Sylvain Munaut)
- rgw: use correct secret key for POST authentication (Robin H. Johnson)

For more detailed information, see [the complete changelog](#).

## v0.67.5 “Dumpling”

---

This release includes a few critical bug fixes for the radosgw, including a fix for hanging operations on large objects. There are also several bug fixes for radosgw multi-site replications, and a few backported features. Also, notably, the ‘osd perf’ command (which dumps recent performance information about active OSDs) has been backported.

We recommend that all 0.67.x Dumpling users upgrade.

## Notable changes

---

- ceph-fuse: fix crash in caching code
- mds: fix looping in populate\_mydir()
- mds: fix standby-replay race
- mon: accept ‘osd pool set ...’ as string
- mon: backport: ‘osd perf’ command to dump recent OSD performance stats
- osd: add feature compat check for upcoming object sharding
- osd: fix osd bench block size argument
- rbd.py: increase parent name size limit
- rgw: backport: allow wildcard in supported keystone roles
- rgw: backport: improve swift COPY behavior
- rgw: backport: log and open admin socket by default
- rgw: backport: validate S3 tokens against keystone
- rgw: fix bucket removal

- rgw: fix client error code for chunked PUT failure
- rgw: fix hang on large object GET
- rgw: fix rare use-after-free
- rgw: various DR bug fixes
- sysvinit, upstart: prevent starting daemons using both init systems

For more detailed information, see [the complete changelog](#).

## v0.67.4 “Dumpling”

---

This point release fixes an important performance issue with radosgw, keystone authentication token caching, and CORS. All users (especially those of rgw) are encouraged to upgrade.

## Notable changes

---

- crush: fix invalidation of cached names
- crushtool: do not crash on non-unique bucket ids
- mds: be more careful when decoding LogEvents
- mds: fix heap check debugging commands
- mon: avoid rebuilding old full osdmaps
- mon: fix ‘ceph crush move ...’
- mon: fix ‘ceph osd crush reweight ...’
- mon: fix writeout of full osdmaps during trim
- mon: limit size of transactions
- mon: prevent both unmanaged and pool snaps
- osd: disable xattr size limit (prevents upload of large rgw objects)
- osd: fix recovery op throttling
- osd: fix throttling of log messages for very slow requests
- rgw: drain pending requests before completing write
- rgw: fix CORS
- rgw: fix inefficient list::size() usage

- rgw: fix keystone token expiration
- rgw: fix minor memory leaks
- rgw: fix null termination of buffer

For more detailed information, see [the complete changelog](#).

## v0.67.3 “Dumpling”

This point release fixes a few important performance regressions with the OSD (both with CPU and disk utilization), as well as several other important but less common problems. We recommend that all production users upgrade.

## Notable Changes

- ceph-disk: partprobe after creation journal partition
- ceph-disk: specify fs type when mounting
- ceph-post-file: new utility to help share logs and other files with ceph developers
- libcephfs: fix truncate vs readahead race (crash)
- mds: fix flock/fcntl lock deadlock
- mds: fix rejoin loop when encountering pre-dumpling backpointers
- mon: allow name and addr discovery during election stage
- mon: always refresh after Paxos store\_state (fixes recovery corner case)
- mon: fix off-by-4x bug with osd byte counts
- osd: add and disable ‘pg log keys debug’ by default
- osd: add option to disable throttling
- osd: avoid leveldb iterators for pg log append and trim
- osd: fix readdir\_r invocations
- osd: use fdatasync instead of sync
- radosgw: fix sysvinit script return status
- rbd: relicense as LGPL2
- rgw: flush pending data on multipart upload

- rgw: recheck object name during S3 POST
- rgw: reorder init/startup
- rpm: fix debuginfo package build

For more detailed information, see [the complete changelog](#).

## v0.67.2 “Dumpling”

---

This is an important point release for Dumpling. Most notably, it fixes a problem when upgrading directly from v0.56.x Bobtail to v0.67.x Dumpling (without stopping at v0.61.x Cuttlefish along the way). It also fixes a problem with the CLI parsing of the CEPH\_ARGS environment variable, high CPU utilization by the ceph-osd daemons, and cleans up the radosgw shutdown sequence.

## Notable Changes

---

- objecter: resend linger requests when cluster goes from full to non-full
- ceph: parse CEPH\_ARGS environment variable
- librados: fix small memory leak
- osd: remove old log objects on upgrade (fixes bobtail -> dumpling jump)
- osd: disable PGLog::check() via config option (fixes CPU burn)
- rgw: drain requests on shutdown
- rgw: misc memory leaks on shutdown

For more detailed information, see [the complete changelog](#).

## v0.67.1 “Dumpling”

---

This is a minor point release for Dumpling that fixes problems with OpenStack and librbd hangs when caching is disabled.

## Notable changes

---

- librados, librbd: fix constructor for python bindings with certain usages (in particular, that used by OpenStack)
- librados, librbd: fix aio\_flush wakeup when cache is disabled
- librados: fix locking for aio completion refcounting

- fixes ‘ceph -admin-daemon ...’ command error code on error
- fixes ‘ceph daemon ... config set ...’ command for boolean config options.

For more detailed information, see [the complete changelog](#).

## v0.67 “Dumpling”

This is the fourth major release of Ceph, code-named “Dumpling.” The headline features for this release include:

- Multi-site support for radosgw. This includes the ability to set up separate “regions” in the same or different Ceph clusters that share a single S3/Swift bucket/container namespace.
- RESTful API endpoint for Ceph cluster administration. ceph-rest-api, a wrapper around ceph\_rest\_api.py, can be used to start up a test single-threaded HTTP server that provides access to cluster information and administration in very similar ways to the ceph commandline tool. ceph\_rest\_api.py can be used as a WSGI application for deployment in a more-capable web server. See ceph-rest-api.8 for more.
- Object namespaces in librados.

## Upgrade Sequencing

It is possible to do a rolling upgrade from Cuttlefish to Dumpling.

1. Upgrade ceph-common on all nodes that will use the command line ‘ceph’ utility.
2. Upgrade all monitors (upgrade ceph package, restart ceph-mon daemons). This can happen one daemon or host at a time. Note that because cuttlefish and dumpling monitors can’t talk to each other, all monitors should be upgraded in relatively short succession to minimize the risk that an untimely failure will reduce availability.
3. Upgrade all osds (upgrade ceph package, restart ceph-osd daemons). This can happen one daemon or host at a time.
4. Upgrade radosgw (upgrade radosgw package, restart radosgw daemons).

## Upgrading from v0.66

- There is monitor internal protocol change, which means that v0.67 ceph-mon daemons cannot talk to v0.66 or older daemons. We recommend upgrading all monitors at once (or in relatively quick succession) to minimize the possibility of downtime.

- The output of ‘ceph status -format=json’ or ‘ceph -s -format=json’ has changed to return status information in a more structured and usable format.
- The ‘ceph pg dump\_stuck [threshold]’ command used to require a -threshold or -t prefix to the threshold argument, but now does not.
- Many more ceph commands now output formatted information; select with ‘-format=<format>’, where <format> can be ‘json’, ‘json-pretty’, ‘xml’, or ‘xml-pretty’.
- The ‘ceph pg <pgid> ...’ commands (like ‘ceph pg <pgid> query’) are deprecated in favor of ‘ceph tell <pgid> ...’. This makes the distinction between ‘ceph pg <command> <pgid>’ and ‘ceph pg <pgid> <command>’ less awkward by making it clearer that the ‘tell’ commands are talking to the OSD serving the placement group, not the monitor.
- The ‘ceph --admin-daemon <path> <command ...>’ used to accept the command and arguments as either a single string or as separate arguments. It will now only accept the command spread across multiple arguments. This means that any script which does something like:

```
1. ceph --admin-daemon /var/run/ceph/ceph-osd.0.asok 'config set debug_ms 1'
```

needs to remove the quotes. Also, note that the above can now be shortened to:

```
1. ceph daemon osd.0 config set debug_ms 1
```

- The radosgw caps were inconsistently documented to be either ‘mon = allow r’ or ‘mon = allow rw’. The ‘mon = allow rw’ is required for radosgw to create its own pools. All documentation has been updated accordingly.
- The radosgw copy object operation may return extra progress info during the operation. At this point it will only happen when doing cross zone copy operations. The S3 response will now return extra <Progress> field under the <CopyResult> container. The Swift response will now send the progress as a json array.
- In v0.66 and v0.65 the HASHPSPPOOL pool flag was enabled by default for new pools, but has been disabled again until Linux kernel client support reaches more distributions and users.
- ceph-osd now requires a max file descriptor limit (e.g., `ulimit -n ...`) of at least `filestore_wbthrottle_(xfs|btrfs)_inodes_hard_limit` (5000 by default) in order to accommodate the new write back throttle system. On Ubuntu, upstart now sets the fd limit to 32k. On other platforms, the sysvinit script will set it to 32k by default (still overrideable via `max_open_files`). If this field has been customized in `ceph.conf` it should likely be adjusted upwards.

# Upgrading from v0.61 “Cuttlefish”

In addition to the above notes about upgrading from v0.66:

- There has been a huge revamp of the ‘ceph’ command-line interface implementation. The `ceph-common` client library needs to be upgrade before `ceph-mon` is restarted in order to avoid problems using the CLI (the old `ceph` client utility cannot talk to the new `ceph-mon` ).
- The CLI is now very careful about sending the ‘status’ one-liner output to stderr and command output to stdout. Scripts relying on output should take care.
- The ‘ceph osd tell ...’ and ‘ceph mon tell ...’ commands are no longer supported. Any callers should use:

```
1. ceph tell osd.<id or *> ...
2. ceph tell mon.<id or name or *> ...
```

The ‘ceph mds tell ...’ command is still there, but will soon also transition to ‘ceph tell mds.<id or name or \*> ...’

- The ‘ceph osd crush add ...’ command used to take one of two forms:

```
1. ceph osd crush add 123 osd.123 <weight> <location ...>
2. ceph osd crush add osd.123 <weight> <location ...>
```

This is because the id and crush name are redundant. Now only the simple form is supported, where the osd name/id can either be a bare id (integer) or name (osd.<id>):

```
1. ceph osd crush add osd.123 <weight> <location ...>
2. ceph osd crush add 123 <weight> <location ...>
```

- There is now a maximum RADOS object size, configurable via ‘osd max object size’, defaulting to 100 GB. Note that this has no effect on RBD, CephFS, or radosgw, which all stripe over objects. If you are using librados and storing objects larger than that, you will need to adjust ‘osd max object size’, and should consider using smaller objects instead.
- The ‘osd min down {reporters|reports}’ config options have been renamed to ‘mon osd min down {reporters|reports}’, and the documentation has been updated to reflect that these options apply to the monitors (who process failure reports) and not OSDs. If you have adjusted these settings, please update your `ceph.conf` accordingly.

## Notable changes since v0.66

- mon: sync improvements (performance and robustness)
- mon: many bug fixes (paxos and services)
- mon: fixed bugs in recovery and io rate reporting (negative/large values)
- mon: collect metadata on osd performance
- mon: generate health warnings from slow or stuck requests
- mon: expanded -format=<json|xml|...> support for monitor commands
- mon: scrub function for verifying data integrity
- mon, osd: fix old osdmap trimming logic
- mon: enable leveldb caching by default
- mon: more efficient storage of PG metadata
- ceph-rest-api: RESTful endpoint for administer cluster (mirrors CLI)
- rgw: multi-region support
- rgw: infrastructure to support georeplication of bucket and user metadata
- rgw: infrastructure to support georeplication of bucket data
- rgw: COPY object support between regions
- rbd: /etc/ceph/rbdmap file for mapping rbd images on startup
- osd: many bug fixes
- osd: limit number of incremental osdmmaps sent to peers (could cause osds to be wrongly marked down)
- osd: more efficient small object recovery
- osd, librados: support for object namespaces
- osd: automatically enable xattrs on leveldb as necessary
- mds: fix bug in LOOKUPINO (used by nfs reexport)
- mds: fix O\_TRUNC locking
- msgr: fixed race condition in inter-osd network communication
- msgr: fixed various memory leaks related to network sessions
- ceph-disk: fixes for unusual device names, partition detection
- hypertable: fixes for hypertable CephBroker bindings

- use SSE4.2 crc32c instruction if present

## Notable changes since v0.61 “Cuttlefish”

---

- add ‘config get’ admin socket command
- ceph-conf: –show-config-value now reflects daemon defaults
- ceph-disk: add ‘[un]suppress-active DEV’ command
- ceph-disk: avoid mounting over an existing osd in /var/lib/ceph/osd/\*
- ceph-disk: fixes for unusual device names, partition detection
- ceph-disk: improved handling of odd device names
- ceph-disk: many fixes for RHEL/CentOS, Fedora, wheezy
- ceph-disk: simpler, more robust locking
- ceph-fuse, libcephfs: fix a few caps revocation bugs
- ceph-fuse, libcephfs: fix read zeroing at EOF
- ceph-fuse, libcephfs: fix request refcounting bug (hang on shutdown)
- ceph-fuse, libcephfs: fix truncatation bug on >4MB files (Yan, Zheng)
- ceph-fuse, libcephfs: fix for cap release/hang
- ceph-fuse: add ioctl support
- ceph-fuse: fixed long-standing O\_NOATIME vs O\_LAZY bug
- ceph-rest-api: RESTful endpoint for administer cluster (mirrors CLI)
- ceph, librados: fix resending of commands on mon reconnect
- daemons: create /var/run/ceph as needed
- debian wheezy: fix udev rules
- debian, specfile: packaging cleanups
- debian: fix upstart behavior with upgrades
- debian: rgw: stop daemon on uninstall
- debian: stop daemons on uninstall; fix dependencies
- hypertable: fixes for hypertable CephBroker bindings
- librados python binding cleanups

- librados python: fix xattrs > 4KB (Josh Durgin)
- librados: configurable max object size (default 100 GB)
- librados: new calls to administer the cluster
- librbd: ability to read from local replicas
- librbd: locking tests (Josh Durgin)
- librbd: make default options/features for newly created images (e.g., via qemu-img) configurable
- librbd: parallelize delete, rollback, flatten, copy, resize
- many many fixes from static code analysis (Danny Al-Gaaf)
- mds: fix O\_TRUNC locking
- mds: fix bug in LOOKUPINO (used by nfs reexport)
- mds: fix rare hang after client restart
- mds: fix several bugs (Yan, Zheng)
- mds: many backpointer improvements (Yan, Zheng)
- mds: many fixes for mds clustering
- mds: misc stability fixes (Yan, Zheng, Greg Farnum)
- mds: new robust open-by-ino support (Yan, Zheng)
- mds: support robust lookup by ino number (good for NFS) (Yan, Zheng)
- mon, ceph: huge revamp of CLI and internal admin API. (Dan Mick)
- mon, osd: fix old osdmap trimming logic
- mon, osd: many memory leaks fixed
- mon: better trim/compaction behavior
- mon: collect metadata on osd performance
- mon: enable leveldb caching by default
- mon: expanded -format=<json|xml|...> support for monitor commands
- mon: fix election timeout
- mon: fix leveldb compression, trimming
- mon: fix start fork behavior

- mon: fix units in ‘ceph df’ output
- mon: fix validation of mds ids from CLI commands
- mon: fixed bugs in recovery and io rate reporting (negative/large values)
- mon: generate health warnings from slow or stuck requests
- mon: many bug fixes (paxos and services, sync)
- mon: many stability fixes (Joao Luis)
- mon: more efficient storage of PG metadata
- mon: new -extract-monmap to aid disaster recovery
- mon: new capability syntax
- mon: scrub function for verifying data integrity
- mon: simplify PaxosService vs Paxos interaction, fix readable/writeable checks
- mon: sync improvements (performance and robustness)
- mon: tuning, performance improvements
- msgr: fix various memory leaks
- msgr: fixed race condition in inter-osd network communication
- msgr: fixed various memory leaks related to network sessions
- osd, librados: support for object namespaces
- osd, mon: optionally dump leveldb transactions to a log
- osd: automatically enable xattrs on leveldb as necessary
- osd: avoid osd flapping from asymmetric network failure
- osd: break blacklisted client watches (David Zafman)
- osd: close narrow journal race
- osd: do not use fadvise(DONTNEED) on XFS (data corruption on power cycle)
- osd: fix for an op ordering bug
- osd: fix handling for split after upgrade from bobtail
- osd: fix incorrect mark-down of osds
- osd: fix internal heartbeat timeouts when scrubbing very large objects

- osd: fix memory/network inefficiency during deep scrub
- osd: fixed problem with front-side heartbeats and mixed clusters (David Zafman)
- osd: limit number of incremental osdmaps sent to peers (could cause osds to be wrongly marked down)
- osd: many bug fixes
- osd: monitor both front and back interfaces
- osd: more efficient small object recovery
- osd: new writeback throttling (for less bursty write performance) (Sam Just)
- osd: pg log (re)writes are now vastly more efficient (faster peering) (Sam Just)
- osd: ping/heartbeat on public and private interfaces
- osd: prioritize recovery for degraded PGs
- osd: re-use partially deleted PG contents when present (Sam Just)
- osd: recovery and peering performance improvements
- osd: resurrect partially deleted PGs
- osd: verify both front and back network are working before rejoining cluster
- rados: clonedata command for cli
- radosgw-admin: create keys for new users by default
- rbd: /etc/ceph/rbdmap file for mapping rbd images on startup
- rgw: COPY object support between regions
- rgw: fix CORS bugs
- rgw: fix locking issue, user operation mask,
- rgw: fix radosgw-admin buckets list (Yehuda Sadeh)
- rgw: fix usage log scanning for large, untrimmed logs
- rgw: handle deep uri resources
- rgw: infrastructure to support georeplication of bucket and user metadata
- rgw: infrastructure to support georeplication of bucket data
- rgw: multi-region support
- sysvinit: fix enumeration of local daemons

- sysvinit: fix osd crush weight calculation when using -a
- sysvinit: handle symlinks in /var/lib/ceph/osd/\*
- use SSE4.2 crc32c instruction if present

## v0.66

---

### Upgrading

- There is now a configurable maximum rados object size, defaulting to 100 GB. If you are using librados and storing objects larger than that, you will need to adjust ‘osd max object size’, and should consider using smaller objects instead.

### Notable changes

- osd: pg log (re)writes are now vastly more efficient (faster peering) (Sam Just)
- osd: fixed problem with front-side heartbeats and mixed clusters (David Zafman)
- mon: tuning, performance improvements
- mon: simplify PaxosService vs Paxos interaction, fix readable/writeable checks
- rgw: fix radosgw-admin buckets list (Yehuda Sadeh)
- mds: support robust lookup by ino number (good for NFS) (Yan, Zheng)
- mds: fix several bugs (Yan, Zheng)
- ceph-fuse, libcephfs: fix truncation bug on >4MB files (Yan, Zheng)
- ceph/librados: fix resending of commands on mon reconnect
- librados python: fix xattrs > 4KB (Josh Durgin)
- librados: configurable max object size (default 100 GB)
- msgr: fix various memory leaks
- ceph-fuse: fixed long-standing O\_NOATIME vs O\_LAZY bug
- ceph-fuse, libcephfs: fix request refcounting bug (hang on shutdown)
- ceph-fuse, libcephfs: fix read zeroing at EOF
- ceph-conf: --show-config-value now reflects daemon defaults
- ceph-disk: simpler, more robust locking

- ceph-disk: avoid mounting over an existing osd in /var/lib/ceph/osd/\*
- sysvinit: handle symlinks in /var/lib/ceph/osd/\*

## v0.65

---

### Upgrading

- Huge revamp of the ‘ceph’ command-line interface implementation. The `ceph-common` client library needs to be upgrade before `ceph-mon` is restarted in order to avoid problems using the CLI (the old `ceph` client utility cannot talk to the new `ceph-mon` ).
- The CLI is now very careful about sending the ‘status’ one-liner output to stderr and command output to stdout. Scripts relying on output should take care.
- The ‘ceph osd tell ...’ and ‘ceph mon tell ...’ commands are no longer supported. Any callers should use:

```
1. ceph tell osd.<id or *> ...
2. ceph tell mon.<id or name or *> ...
```

The ‘ceph mds tell ...’ command is still there, but will soon also transition to ‘ceph tell mds.<id or name or \*> ...’

- The ‘ceph osd crush add ...’ command used to take one of two forms:

```
1. ceph osd crush add 123 osd.123 <weight> <location ...>
2. ceph osd crush add osd.123 <weight> <location ...>
```

This is because the id and crush name are redundant. Now only the simple form is supported, where the osd name/id can either be a bare id (integer) or name (osd.<id>):

```
1. ceph osd crush add osd.123 <weight> <location ...>
2. ceph osd crush add 123 <weight> <location ...>
```

- There is now a maximum RADOS object size, configurable via ‘osd max object size’, defaulting to 100 GB. Note that this has no effect on RBD, CephFS, or radosgw, which all stripe over objects.

### Notable changes

---

- mon, ceph: huge revamp of CLI and internal admin API. (Dan Mick)

- mon: new capability syntax
- osd: do not use fadvise(DONTNEED) on XFS (data corruption on power cycle)
- osd: recovery and peering performance improvements
- osd: new writeback throttling (for less bursty write performance) (Sam Just)
- osd: ping/heartbeat on public and private interfaces
- osd: avoid osd flapping from asymmetric network failure
- osd: re-use partially deleted PG contents when present (Sam Just)
- osd: break blacklisted client watches (David Zafman)
- mon: many stability fixes (Joao Luis)
- mon, osd: many memory leaks fixed
- mds: misc stability fixes (Yan, Zheng, Greg Farnum)
- mds: many backpointer improvements (Yan, Zheng)
- mds: new robust open-by-ino support (Yan, Zheng)
- ceph-fuse, libcephfs: fix a few caps revocation bugs
- librados: new calls to administer the cluster
- librbd: locking tests (Josh Durgin)
- ceph-disk: improved handling of odd device names
- ceph-disk: many fixes for RHEL/CentOS, Fedora, wheezy
- many many fixes from static code analysis (Danny Al-Gaaf)
- daemons: create /var/run/ceph as needed

## v0.64

---

### Upgrading

---

- New pools now have the HASHPSPPOOL flag set by default to provide better distribution over OSDs. Support for this feature was introduced in v0.59 and Linux kernel version v3.9. If you wish to access the cluster from an older kernel, set the ‘osd pool default flag hashpspool = false’ option in your ceph.conf prior to creating the cluster or creating new pools. Note that the presence of any pool in the cluster with the flag enabled will make the OSD

require support from all clients.

## Notable changes

---

- osd: monitor both front and back interfaces
- osd: verify both front and back network are working before rejoining cluster
- osd: fix memory/network inefficiency during deep scrub
- osd: fix incorrect mark-down of osds
- mon: fix start fork behavior
- mon: fix election timeout
- mon: better trim/compaction behavior
- mon: fix units in ‘ceph df’ output
- mon, osd: misc memory leaks
- librbd: make default options/features for newly created images (e.g., via qemu-img) configurable
- mds: many fixes for mds clustering
- mds: fix rare hang after client restart
- ceph-fuse: add ioctl support
- ceph-fuse/libcephfs: fix for cap release/hang
- rgw: handle deep uri resources
- rgw: fix CORS bugs
- ceph-disk: add ‘[un]suppress-active DEV’ command
- debian: rgw: stop daemon on uninstall
- debian: fix upstart behavior with upgrades

## v0.63

---

## Upgrading

---

- The ‘osd min down {reporters|reports}’ config options have been renamed to ‘mon osd min down {reporters|reports}’, and the documentation has been updated to

reflect that these options apply to the monitors (who process failure reports) and not OSDs. If you have adjusted these settings, please update your `ceph.conf` accordingly.

## Notable Changes

---

- librbd: parallelize delete, rollback, flatten, copy, resize
- librbd: ability to read from local replicas
- osd: resurrect partially deleted PGs
- osd: prioritize recovery for degraded PGs
- osd: fix internal heartbeat timeouts when scrubbing very large objects
- osd: close narrow journal race
- rgw: fix usage log scanning for large, untrimmed logs
- rgw: fix locking issue, user operation mask,
- initscript: fix osd crush weight calculation when using -a
- initscript: fix enumeration of local daemons
- mon: several fixes to paxos, sync
- mon: new -extract-monmap to aid disaster recovery
- mon: fix leveldb compression, trimming
- add ‘config get’ admin socket command
- rados: clonedata command for cli
- debian: stop daemons on uninstall; fix dependencies
- debian wheezy: fix udev rules
- many many small fixes from coverity scan

## v0.62

---

## Notable Changes

---

- mon: fix validation of mds ids from CLI commands
- osd: fix for an op ordering bug

- osd, mon: optionally dump leveldb transactions to a log
- osd: fix handling for split after upgrade from bobtail
- debian, specfile: packaging cleanups
- radosgw-admin: create keys for new users by default
- librados python binding cleanups
- misc code cleanups

# v0.61.9 “Cuttlefish”

This point release resolves several low to medium-impact bugs across the code base, and fixes a performance problem (CPU utilization) with radosgw. We recommend that all production cuttlefish users upgrade.

## Notable Changes

- ceph, ceph-authtool: fix help (Danny Al-Gaaf)
- ceph-disk: partprobe after creating journal partition
- ceph-disk: specific fs type when mounting (Alfredo Deza)
- ceph-fuse: fix bug when compiled against old versions
- ceph-fuse: fix use-after-free in caching code (Yan, Zheng)
- ceph-fuse: misc caching bugs
- ceph.spec: remove incorrect mod\_fcgi dependency (Gary Lowell)
- crush: fix name caching
- librbd: fix bug when unpausing cluster (Josh Durgin)
- mds: fix LAZYIO lock hang
- mds: fix bug in file size recovery (after client crash)
- mon: fix paxos recovery corner case
- osd: fix exponential backoff for slow request warnings (Loic Dachary)
- osd: fix readdir\_r usage
- osd: fix startup for long-stopped OSDs
- rgw: avoid std::list::size() to avoid wasting CPU cycles (Yehuda Sadeh)
- rgw: drain pending requests during write (fixes data safety issue) (Yehuda Sadeh)
- rgw: fix authenticated users group ACL check (Yehuda Sadeh)
- rgw: fix bug in POST (Yehuda Sadeh)
- rgw: fix sysvinit script ‘status’ command, return value (Danny Al-Gaaf)
- rgw: reduce default log level (Yehuda Sadeh)

For more detailed information, see [the complete changelog](#).

## v0.61.8 “Cuttlefish”

This release includes a number of important issues, including rare race conditions in the OSD, a few monitor bugs, and fixes for RBD flush behavior. We recommend that production users upgrade at their convenience.

## Notable Changes

- librados: fix async aio completion wakeup
- librados: fix aio completion locking
- librados: fix rare deadlock during shutdown
- osd: fix race when queueing recovery operations
- osd: fix possible race during recovery
- osd: optionally preload rados classes on startup (disabled by default)
- osd: fix journal replay corner condition
- osd: limit size of peering work queue batch (to speed up peering)
- mon: fix paxos recovery corner case
- mon: fix rare hang when monmap updates during an election
- mon: make ‘osd pool mksnap ...’ avoid exposing uncommitted state
- mon: make ‘osd pool rmsnap ...’ not racy, avoid exposing uncommitted state
- mon: fix bug during mon cluster expansion
- rgw: fix crash during multi delete operation
- msgr: fix race conditions during osd network reinitialization
- ceph-disk: apply mount options when remounting

For more detailed information, see [the complete changelog](#).

## v0.61.7 “Cuttlefish”

This release fixes another regression preventing monitors to start after undergoing certain upgrade sequences, as well as some corner cases with Paxos and support for unusual device names in ceph-disk/ceph-deploy.

## Notable Changes

---

- mon: fix regression in latest full osdmap retrieval
- mon: fix a long-standing bug in a paxos corner case
- ceph-disk: improved support for unusual device names (e.g., /dev/cciss/c0d0)

For more detailed information, see [the complete changelog](#).

## v0.61.6 “Cuttlefish”

---

This release fixes a regression in v0.61.5 that could prevent monitors from restarting. This affects any cluster that was upgraded from a previous version of Ceph (and not freshly created with v0.61.5).

All users are strongly recommended to upgrade.

## Notable Changes

---

- mon: record latest full osdmap
- mon: work around previous bug in which latest full osdmap is not recorded
- mon: avoid scrub while updating

For more detailed information, see [the complete changelog](#).

## v0.61.5 “Cuttlefish”

---

This release most improves stability of the monitor and fixes a few bugs with the ceph-disk utility (used by ceph-deploy). We recommend that all v0.61.x users upgrade.

## Upgrading

---

- This release fixes a 32-bit vs 64-bit arithmetic bug with the feature bits. An unfortunate consequence of the fix is that 0.61.4 (or earlier) ceph-mon daemons can't form a quorum with 0.61.5 (or later) monitors. To avoid the possibility of service disruption, we recommend you upgrade all monitors at once.

## Notable Changes

---

- mon: misc sync improvements (faster, more reliable, better tuning)
- mon: enable leveldb cache by default (big performance improvement)

- mon: new scrub feature (primarily for diagnostic, testing purposes)
- mon: fix occasional leveldb assertion on startup
- mon: prevent reads until initial state is committed
- mon: improved logic for trimming old osdmaps
- mon: fix pick\_addresses bug when expanding mon cluster
- mon: several small paxos fixes, improvements
- mon: fix bug osdmap trim behavior
- osd: fix several bugs with PG stat reporting
- osd: limit number of maps shared with peers (which could cause domino failures)
- rgw: fix radosgw-admin buckets list (for all buckets)
- mds: fix occasional client failure to reconnect
- mds: fix bad list traversal after unlink
- mds: fix underwater dentry cleanup (occasional crash after mds restart)
- libcephfs, ceph-fuse: fix occasional hangs on umount
- libcephfs, ceph-fuse: fix old bug with O\_LAZY vs O\_NOATIME confusion
- ceph-disk: more robust journal device detection on RHEL/CentOS
- ceph-disk: better, simpler locking
- ceph-disk: do not inadvertently mount over existing osd mounts
- ceph-disk: better handling for unusual device names
- sysvinit, upstart: handle symlinks in /var/lib/ceph/\*

For more detailed information, see [the complete changelog](#).

## v0.61.4 “Cuttlefish”

---

This release resolves a possible data corruption on power-cycle when using XFS, a few outstanding problems with monitor sync, several problems with ceph-disk and ceph-deploy operation, and a problem with OSD memory usage during scrub.

## Upgrading

---

- No issues.

## Notable Changes

- mon: fix daemon exit behavior when error is encountered on startup
- mon: more robust sync behavior
- osd: do not use sync\_file\_range(2), posix\_fadvise(...DONTNEED) (can cause data corruption on power loss on XFS)
- osd: avoid unnecessary log rewrite (improves peering speed)
- osd: fix scrub efficiency bug (problematic on old clusters)
- rgw: fix listing objects that start with underscore
- rgw: fix deep URI resource, CORS bugs
- librados python binding: fix truncate on 32-bit architectures
- ceph-disk: fix udev rules
- rpm: install sysvinit script on package install
- ceph-disk: fix OSD start on machine reboot on Debian wheezy
- ceph-disk: activate OSD when journal device appears second
- ceph-disk: fix various bugs on RHEL/CentOS 6.3
- ceph-disk: add ‘zap’ command
- ceph-disk: add ‘[un]suppress-activate’ command for preparing spare disks
- upstart: start on runlevel [2345] (instead of after the first network interface starts)
- ceph-fuse, libcephfs: handle mds session reset during session open
- ceph-fuse, libcephfs: fix two capability revocation bugs
- ceph-fuse: fix thread creation on startup
- all daemons: create /var/run/ceph directory on startup if missing

For more detailed information, see [the complete changelog](#).

## v0.61.3 “Cuttlefish”

This release resolves a number of problems with the monitors and leveldb that users have been seeing. Please upgrade.

# Upgrading

---

- There is one known problem with mon upgrades from bobtail. If the ceph-mon conversion on startup is aborted or fails for some reason, we do not correctly error out, but instead continue with (in certain cases) odd results. Please be careful if you have to restart the mons during the upgrade. A 0.61.4 release with a fix will be out shortly.
- In the meantime, for current cuttlefish users, v0.61.3 is safe to use.

## Notable Changes

---

- mon: paxos state trimming fix (resolves runaway disk usage)
- mon: finer-grained compaction on trim
- mon: discard messages from disconnected clients (lowers load)
- mon: leveldb compaction and other stats available via admin socket
- mon: async compaction (lower overhead)
- mon: fix bug incorrectly marking osds down with insufficient failure reports
- osd: fixed small bug in pg request map
- osd: avoid rewriting pg info on every osdmap
- osd: avoid internal heartbeta timeouts when scrubbing very large objects
- osd: fix narrow race with journal replay
- mon: fixed narrow pg split race
- rgw: fix leaked space when copying object
- rgw: fix iteration over large/untrimmed usage logs
- rgw: fix locking issue with ops log socket
- rgw: require matching version of librados
- librbd: make image creation defaults configurable (e.g., create format 2 images via qemu-img)
- fix units in ‘ceph df’ output
- debian: fix prerm/postinst hooks to start/stop daemons appropriately
- upstart: allow uppercase daemons names (and thus hostnames)

- sysvinit: fix enumeration of local daemons by type
- sysvinit: fix osd weight calcuation when using -a
- fix build on unsigned char platforms (e.g., arm)

For more detailed information, see [the complete changelog](#).

## v0.61.2 “Cuttlefish”

---

This release disables a monitor debug log that consumes disk space and fixes a bug when upgrade some monitors from bobtail to cuttlefish.

### Notable Changes

---

- mon: fix conversion of stores with duplicated GV values
- mon: disable ‘mon debug dump transactions’ by default

For more detailed information, see [the complete changelog](#).

## v0.61.1 “Cuttlefish”

---

This release fixes a problem when upgrading a bobtail cluster that had snapshots to cuttlefish.

### Notable Changes

---

- osd: handle upgrade when legacy snap collections are present; repair from previous failed restart
- ceph-create-keys: fix race with ceph-mon startup (which broke ‘ceph-deploy gatherkeys ...’)
- ceph-create-keys: gracefully handle bad response from ceph-osd
- sysvinit: do not assume default osd\_data when automatically weighting OSD
- osd: avoid crash from ill-behaved classes using getomapvals
- debian: fix squeeze dependency
- mon: debug options to log or dump leveldb transactions

For more detailed information, see [the complete changelog](#).

# v0.61 “Cuttlefish”

---

## Upgrading from v0.60

- The ceph-deploy tool is now the preferred method of provisioning new clusters. For existing clusters created via mkcephfs that would like to transition to the new tool, there is a migration path, documented at [Transitioning to ceph-deploy](#).
- The sysvinit script (/etc/init.d/ceph) will now verify (and, if necessary, update) the OSD’s position in the CRUSH map on startup. (The upstart script has always worked this way.) By default, this ensures that the OSD is under a ‘host’ with a name that matches the hostname (`hostname -s`). Legacy clusters create with mkcephfs do this by default, so this should not cause any problems, but legacy clusters with customized CRUSH maps with an alternate structure should set `osd crush update on start = false`.
- radosgw-admin now uses the term zone instead of cluster to describe each instance of the radosgw data store (and corresponding collection of radosgw daemons). The usage for the radosgw-admin command and the ‘rgw zone root pool’ config options have changed accordingly.
- rbd progress indicators now go to standard error instead of standard out. (You can disable progress with `-no-progress`.)
- The ‘rbd resize ...’ command now requires the `-allow-shrink` option when resizing to a smaller size. Expanding images to a larger size is unchanged.
- Please review the changes going back to 0.56.4 if you are upgrading all the way from bobtail.
- The old ‘ceph stop\_cluster’ command has been removed.
- The sysvinit script now uses the `ceph.conf` file on the remote host when starting remote daemons via the ‘`-a`’ option. Note that if ‘`-a`’ is used in conjunction with ‘`-c path`’, the path must also be present on the remote host (it is not copied to a temporary file, as it was previously).

## Upgrading from v0.56.4 “Bobtail”

---

Please see [Upgrading from Bobtail to Cuttlefish](#) for details.

- The ceph-deploy tool is now the preferred method of provisioning new clusters. For existing clusters created via mkcephfs that would like to transition to the new tool, there is a migration path, documented at [Transitioning to ceph-deploy](#).
- The sysvinit script (/etc/init.d/ceph) will now verify (and, if necessary, update) the OSD’s position in the CRUSH map on startup. (The upstart script has

always worked this way.) By default, this ensures that the OSD is under a ‘host’ with a name that matches the hostname (`hostname -s`). Legacy clusters create with `mkcephfs` do this by default, so this should not cause any problems, but legacy clusters with customized CRUSH maps with an alternate structure should set `osd crush update on start = false`.

- `radosgw-admin` now uses the term `zone` instead of `cluster` to describe each instance of the `radosgw` data store (and corresponding collection of `radosgw` daemons). The usage for the `radosgw-admin` command and the ‘`rgw zone root pool`’ config options have changed accordingly.
- `rbd` progress indicators now go to standard error instead of standard out. (You can disable progress with `-no-progress`.)
- The ‘`rbd resize ...`’ command now requires the `-allow-shrink` option when resizing to a smaller size. Expanding images to a larger size is unchanged.
- Please review the changes going back to 0.56.4 if you are upgrading all the way from `bobtail`.
- The old ‘`ceph stop_cluster`’ command has been removed.
- The `sysvinit` script now uses the `ceph.conf` file on the remote host when starting remote daemons via the ‘`-a`’ option. Note that if ‘`-a`’ is used in conjunction with ‘`-c path`’, the path must also be present on the remote host (it is not copied to a temporary file, as it was previously).
- The monitor is using a completely new storage strategy and intra-cluster protocol. This means that `cuttlefish` and `bobtail` monitors do not talk to each other. When you upgrade each one, it will convert its local data store to the new format. Once you upgrade a majority, the quorum will be formed using the new protocol and the old monitors will be blocked out until they too get upgraded. For this reason, we recommend not running a mixed-version cluster for very long.
- `ceph-mon` now requires the creation of its data directory prior to `-mkfs`, similarly to what happens on `ceph-osd`. This directory is no longer automatically created, and custom scripts should be adjusted to reflect just that.
- The monitor now enforces that MDS names be unique. If you have multiple daemons start with the same id (e.g., `mds.a`) the second one will implicitly mark the first as failed. This makes things less confusing and makes a daemon restart faster (we no longer wait for the stopped daemon to time out) but existing multi-mds configurations may need to be adjusted accordingly to give daemons unique names.
- The ‘`ceph osd pool delete <poolname>`’ and ‘`rados rm pool <poolname>`’ now have safety interlocks with loud warnings that make you confirm pool removal. Any scripts currently rely on these functions zapping data without confirmation need to be adjusted accordingly.

# Notable Changes from v0.60

---

- rbd: incremental backups
- rbd: only set STRIPEVG2 feature if striping parameters are incompatible with old versions
- rbd: require -allow-shrink for resizing images down
- librbd: many bug fixes
- rgw: management REST API
- rgw: fix object corruption on COPY to self
- rgw: new sysvinit script for rpm-based systems
- rgw: allow buckets with ‘\_’
- rgw: CORS support
- mon: many fixes
- mon: improved trimming behavior
- mon: fix data conversion/upgrade problem (from bobtail)
- mon: ability to tune leveldb
- mon: config-keys service to store arbitrary data on monitor
- mon: ‘osd crush add|link|unlink|add-bucket ...’ commands
- mon: trigger leveldb compaction on trim
- osd: per-rados pool quotas (objects, bytes)
- osd: tool to export, import, and delete PGs from an individual OSD data store
- osd: notify mon on clean shutdown to avoid IO stall
- osd: improved detection of corrupted journals
- osd: ability to tune leveldb
- osd: improve client request throttling
- osd, librados: fixes to the LIST\_SNAPS operation
- osd: improvements to scrub error repair
- osd: better prevention of wedging OSDs with ENOSPC

- osd: many small fixes
- mds: fix xattr handling on root inode
- mds: fixed bugs in journal replay
- mds: many fixes
- librados: clean up snapshot constant definitions
- libcephfs: calls to query CRUSH topology (used by Hadoop)
- ceph-fuse, libcephfs: misc fixes to mds session management
- ceph-fuse: disabled cache invalidation (again) due to potential deadlock with kernel
- sysvinit: try to start all daemons despite early failures
- ceph-disk: new ‘list’ command
- ceph-disk: hotplug fixes for RHEL/CentOS
- ceph-disk: fix creation of OSD data partitions on >2TB disks
- osd: fix udev rules for RHEL/CentOS systems
- fix daemon logging during initial startup

## Notable changes from v0.56 “Bobtail”

---

- always use installed system leveldb (Gary Lowell)
- auth: ability to require new cephx signatures on messages (still off by default)
- buffer unit testing (Loic Dachary)
- ceph tool: some CLI interface cleanups
- ceph-disk: improve multicluster support, error handling (Sage Weil)
- ceph-disk: support for dm-crypt (Alexandre Marangone)
- ceph-disk: support for sysvinit, directories or partitions (not full disks)
- ceph-disk: fix mkfs args on old distros (Alexandre Marangone)
- ceph-disk: fix creation of OSD data partitions on >2TB disks
- ceph-disk: hotplug fixes for RHEL/CentOS
- ceph-disk: new ‘list’ command

- ceph-fuse, libcephfs: misc fixes to mds session management
- ceph-fuse: disabled cache invalidation (again) due to potential deadlock with kernel
- ceph-fuse: enable kernel cache invalidation (Sam Lang)
- ceph-fuse: fix statfs(2) reporting
- ceph-fuse: session handling cleanup, bug fixes (Sage Weil)
- crush: ability to create, remove rules via CLI
- crush: update weights for all instances of an item, not just the first (Sage Weil)
- fix daemon logging during initial startup
- fixed log rotation (Gary Lowell)
- init-ceph, mkcephfs: close a few security holes with -a (Sage Weil)
- libcephfs: calls to query CRUSH topology (used by Hadoop)
- libcephfs: many fixes, cleanups with the Java bindings
- libcephfs: new topo API requests for Hadoop (Noah Watkins)
- librados: clean up snapshot constant definitions
- librados: fix linger bugs (Josh Durgin)
- librbd: fixed flatten deadlock (Josh Durgin)
- librbd: fixed some locking issues with flatten (Josh Durgin)
- librbd: many bug fixes
- librbd: optionally wait for flush before enabling writeback (Josh Durgin)
- many many cleanups (Danny Al-Gaaf)
- mds, ceph-fuse: fix bugs with replayed requests after MDS restart (Sage Weil)
- mds, ceph-fuse: manage layouts via xattrs
- mds: allow xattrs on root
- mds: fast failover between MDSs (enforce unique mds names)
- mds: fix xattr handling on root inode
- mds: fixed bugs in journal replay

- mds: improve session cleanup (Sage Weil)
- mds: many fixes (Yan Zheng)
- mds: misc bug fixes with clustered MDSS and failure recovery
- mds: misc bug fixes with readdir
- mds: new encoding for all data types (to allow forward/backward compatibility) (Greg Farnum)
- mds: store and update backpointers/traces on directory, file objects (Sam Lang)
- mon: ‘osd crush add|link|unlink|add-bucket ...’ commands
- mon: ability to tune leveldb
- mon: approximate recovery, IO workload stats
- mon: avoid marking entire CRUSH subtrees out (e.g., if an entire rack goes offline)
- mon: config-keys service to store arbitrary data on monitor
- mon: easy adjustment of crush tunables via ‘ceph osd crush tunables ...’
- mon: easy creation of crush rules via ‘ceph osd rule ...’
- mon: fix data conversion/upgrade problem (from bobtail)
- mon: improved trimming behavior
- mon: many fixes
- mon: new ‘ceph df [detail]’ command
- mon: new checks for identifying and reporting clock drift
- mon: rearchitected to utilize single instance of paxos and a key/value store (Joao Luis)
- mon: safety check for pool deletion
- mon: shut down safely if disk approaches full (Joao Luis)
- mon: trigger leveldb compaction on trim
- msgr: fix comparison of IPv6 addresses (fixes monitor bringup via ceph-deploy, chef)
- msgr: fixed race in connection reset
- msgr: optionally tune TCP buffer size to avoid throughput collapse (Jim Schutt)

- much code cleanup and optimization (Danny Al-Gaaf)
- osd, librados: ability to list watchers (David Zafman)
- osd, librados: fixes to the LIST\_SNAPS operation
- osd, librados: new listsnaps command (David Zafman)
- osd: a few journaling bug fixes
- osd: ability to tune leveldb
- osd: add ‘noscrub’, ‘nodeepscrub’ osdmap flags (David Zafman)
- osd: better prevention of wedging OSDs with ENOSPC
- osd: ceph-filestore-dump tool for debugging
- osd: connection handling bug fixes
- osd: deep-scrub omap keys/values
- osd: default to libaio for the journal (some performance boost)
- osd: fix hang in ‘journal aio = true’ mode (Sage Weil)
- osd: fix pg log trimming (avoids memory bloat on degraded clusters)
- osd: fix udev rules for RHEL/CentOS systems
- osd: fixed bug in journal checksums (Sam Just)
- osd: improved client request throttling
- osd: improved handling when disk fills up (David Zafman)
- osd: improved journal corruption detection (Sam Just)
- osd: improved detection of corrupted journals
- osd: improvements to scrub error repair
- osd: make tracking of object snapshot metadata more efficient (Sam Just)
- osd: many small fixes
- osd: misc fixes to PG split (Sam Just)
- osd: move pg info, log into leveldb (== better performance) (David Zafman)
- osd: notify mon on clean shutdown to avoid IO stall
- osd: per-rados pool quotas (objects, bytes)

- osd: refactored watch/notify infrastructure (fixes protocol, removes many bugs) (Sam Just)
- osd: support for improved hashing of PGs across OSDs via HASHPSPOOL pool flag and feature
- osd: tool to export, import, and delete PGs from an individual OSD data store
- osd: trim log more aggressively, avoid appearance of leak memory
- osd: validate snap collections on startup
- osd: verify snap collections on startup (Sam Just)
- radosgw: ACL grants in headers (Caleb Miles)
- radosgw: ability to listen to fastcgi via a port (Guilhem Lettron)
- radosgw: fix object copy onto self (Yehuda Sadeh)
- radosgw: misc fixes
- rbd-fuse: new tool, package
- rbd: avoid FIEMAP when importing from file (it can be buggy)
- rbd: incremental backups
- rbd: only set STRIPINGV2 feature if striping parameters are incompatible with old versions
- rbd: require -allow-shrink for resizing images down
- rbd: udevadm settle on map/unmap to avoid various races (Dan Mick)
- rbd: wait for udev to settle in strategic places (avoid spurious errors, failures)
- rgw: CORS support
- rgw: allow buckets with ‘\_’
- rgw: fix Content-Length on 32-bit machines (Jan Harkes)
- rgw: fix log rotation
- rgw: fix object corruption on COPY to self
- rgw: fixed >4MB range requests (Jan Harkes)
- rgw: new sysvinit script for rpm-based systems
- rpm/deb: do not remove /var/lib/ceph on purge (v0.59 was the only release to do

so)

- sysvinit: try to start all daemons despite early failures
- upstart: automatically set osd weight based on df (Guilhem Lettron)
- use less memory for logging by default

## v0.60

---

### Upgrading

- Please note that the recently added librados ‘list\_snaps’ function call is in a state of flux and is changing slightly in v0.61. You are advised not to make use of it in v0.59 or v0.60.

### Notable Changes

- osd: make tracking of object snapshot metadata more efficient (Sam Just)
- osd: misc fixes to PG split (Sam Just)
- osd: improve journal corruption detection (Sam Just)
- osd: improve handling when disk fills up (David Zafman)
- osd: add ‘noscrub’, ‘nodeepscrub’ osdmap flags (David Zafman)
- osd: fix hang in ‘journal aio = true’ mode (Sage Weil)
- ceph-disk-prepare: fix mkfs args on old distros (Alexandre Marangone)
- ceph-disk-activate: improve multicluster support, error handling (Sage Weil)
- librbd: optionally wait for flush before enabling writeback (Josh Durgin)
- crush: update weights for all instances of an item, not just the first (Sage Weil)
- mon: shut down safely if disk approaches full (Joao Luis)
- rgw: fix Content-Length on 32-bit machines (Jan Harkes)
- mds: store and update backpointers/traces on directory, file objects (Sam Lang)
- mds: improve session cleanup (Sage Weil)
- mds, ceph-fuse: fix bugs with replayed requests after MDS restart (Sage Weil)

- ceph-fuse: enable kernel cache invalidation (Sam Lang)
- libcephfs: new topo API requests for Hadoop (Noah Watkins)
- ceph-fuse: session handling cleanup, bug fixes (Sage Weil)
- much code cleanup and optimization (Danny Al-Gaaf)
- use less memory for logging by default
- upstart: automatically set osd weight based on df (Guilhem Lettron)
- init-ceph, mkcephfs: close a few security holes with -a (Sage Weil)
- rpm/deb: do not remove /var/lib/ceph on purge (v0.59 was the only release to do so)

## v0.59

---

### Upgrading

---

- The monitor is using a completely new storage strategy and intra-cluster protocol. This means that v0.59 and pre-v0.59 monitors do not talk to each other. When you upgrade each one, it will convert its local data store to the new format. Once you upgrade a majority, the quorum will be formed using the new protocol and the old monitors will be blocked out until they too get upgraded. For this reason, we recommend not running a mixed-version cluster for very long.
- ceph-mon now requires the creation of its data directory prior to -mkfs, similarly to what happens on ceph-osd. This directory is no longer automatically created, and custom scripts should be adjusted to reflect just that.

### Notable Changes

---

- mon: rearchitected to utilize single instance of paxos and a key/value store (Joao Luis)
  - mon: new ‘ceph df [detail]’ command
  - osd: support for improved hashing of PGs across OSDs via HASHPSPOOL pool flag and feature
  - osd: refactored watch/notify infrastructure (fixes protocol, removes many bugs) (Sam Just)
  - osd, librados: ability to list watchers (David Zafman)
  - osd, librados: new listsnaps command (David Zafman)
  - osd: trim log more aggressively, avoid appearance of leak memory
  - osd: misc split fixes

- osd: a few journaling bug fixes
- osd: connection handling bug fixes
- rbd: avoid FIEMAP when importing from file (it can be buggy)
- librados: fix linger bugs (Josh Durgin)
- librbd: fixed flatten deadlock (Josh Durgin)
- rgw: fixed >4MB range requests (Jan Harkes)
- rgw: fix log rotation
- mds: allow xattrs on root
- ceph-fuse: fix statfs(2) reporting
- msgr: optionally tune TCP buffer size to avoid throughput collapse (Jim Schutt)
- consume less memory for logging by default
- always use system leveldb (Gary Lowell)

## v0.58

---

### Upgrading

- The monitor now enforces that MDS names be unique. If you have multiple daemons start with the same id (e.g., `mds.a`) the second one will implicitly mark the first as failed. This makes things less confusing and makes a daemon restart faster (we no longer wait for the stopped daemon to time out) but existing multi-mds configurations may need to be adjusted accordingly to give daemons unique names.

### Notable Changes

---

- librbd: fixed some locking issues with flatten (Josh Durgin)
- rbd: udevadm settle on map/unmap to avoid various races (Dan Mick)
- osd: move pg info, log into leveldb (== better performance) (David Zafman)
- osd: fix pg log trimming (avoids memory bloat on degraded clusters)
- osd: fixed bug in journal checksums (Sam Just)
- osd: verify snap collections on startup (Sam Just)
- ceph-disk-prepare/activate: support for dm-crypt (Alexandre Marangone)
- ceph-disk-prepare/activate: support for sysvinit, directories or partitions (not full disks)

- msgr: fixed race in connection reset
- msgr: fix comparison of IPv6 addresses (fixes monitor bringup via ceph-deploy, chef)
- radosgw: fix object copy onto self (Yehuda Sadeh)
- radosgw: ACL grants in headers (Caleb Miles)
- radosgw: ability to listen to fastcgi via a port (Guilhem Lettron)
- mds: new encoding for all data types (to allow forward/backward compatibility) (Greg Farnum)
- mds: fast failover between MDSs (enforce unique mds names)
- crush: ability to create, remove rules via CLI
- many many cleanups (Danny Al-Gaaf)
- buffer unit testing (Loic Dachary)
- fixed log rotation (Gary Lowell)

## v0.57

This development release has a lot of additional functionality accumulated over the last couple months. Most of the bug fixes (with the notable exception of the MDS related work) has already been backported to v0.56.x, and is not mentioned here.

## Upgrading

- The ‘ceph osd pool delete <poolname>’ and ‘rados rm pool <poolname>’ now have safety interlocks with loud warnings that make you confirm pool removal. Any scripts currently rely on these functions zapping data without confirmation need to be adjusted accordingly.

## Notable Changes

- osd: default to libaio for the journal (some performance boost)
- osd: validate snap collections on startup
- osd: ceph-filestore-dump tool for debugging
- osd: deep-scrub omap keys/values
- ceph tool: some CLI interface cleanups
- mon: easy adjustment of crush tunables via ‘ceph osd crush tunables ...’
- mon: easy creation of crush rules via ‘ceph osd rule ...’

- mon: approximate recovery, IO workload stats
- mon: avoid marking entire CRUSH subtrees out (e.g., if an entire rack goes offline)
- mon: safety check for pool deletion
- mon: new checks for identifying and reporting clock drift
- radosgw: misc fixes
- rbd: wait for udev to settle in strategic places (avoid spurious errors, failures)
- rbd-fuse: new tool, package
- mds, ceph-fuse: manage layouts via xattrs
- mds: misc bug fixes with clustered MDSs and failure recovery
- mds: misc bug fixes with readdir
- libcephfs: many fixes, cleanups with the Java bindings
- auth: ability to require new cephx signatures on messages (still off by default)

# v0.56.7 “bobtail”

This bobtail update fixes a range of radosgw bugs (including an easily triggered crash from multi-delete), a possible data corruption issue with power failure on XFS, and several OSD problems, including a memory “leak” that will affect aged clusters.

## Notable changes

- ceph-fuse: create finisher flags after fork()
- debian: fix prerm/postinst hooks; do not restart daemons on upgrade
- librados: fix async aio completion wakeup (manifests as rbd hang)
- librados: fix hang when osd becomes full and then not full
- librados: fix locking for aio completion refcounting
- librbd python bindings: fix stripe\_unit, stripe\_count
- librbd: make image creation default configurable
- mon: fix validation of mds ids in mon commands
- osd: avoid excessive disk updates during peering
- osd: avoid excessive memory usage on scrub
- osd: avoid heartbeat failure/suicide when scrubbing
- osd: misc minor bug fixes
- osd: use fdatasync instead of sync\_file\_range (may avoid xfs power-loss corruption)
- rgw: escape prefix correctly when listing objects
- rgw: fix copy attrs
- rgw: fix crash on multi delete
- rgw: fix locking/crash when using ops log socket
- rgw: fix usage logging
- rgw: handle deep uri resources

For more detailed information, see [the complete changelog](#).

## v0.56.6 “bobtail”

### Notable changes

- rgw: fix garbage collection
- rpm: fix package dependencies

For more detailed information, see [the complete changelog](#).

## v0.56.5 “bobtail”

### Upgrading

- ceph-disk[-prepare,-activate] behavior has changed in various ways. There should not be any compatibility issues, but chef users should be aware.

### Notable changes

- mon: fix recording of quorum feature set (important for argonaut -> bobtail -> cuttlefish mon upgrades)
- osd: minor peering bug fixes
- osd: fix a few bugs when pools are renamed
- osd: fix occasionally corrupted pg stats
- osd: fix behavior when broken v0.56[.0] clients connect
- rbd: avoid FIEMAP ioctl on import (it is broken on some kernels)
- librbd: fixes for several request/reply ordering bugs
- librbd: only set STRIPINGV2 feature on new images when needed
- librbd: new async flush method to resolve qemu hangs (requires QEMU update as well)
- librbd: a few fixes to flatten
- ceph-disk: support for dm-crypt
- ceph-disk: many backports to allow bobtail deployments with ceph-deploy, chef
- sysvinit: do not stop starting daemons on first failure

- udev: fixed rules for redhat-based distros
- build fixes for raring

For more detailed information, see [the complete changelog](#).

## v0.56.4 “bobtail”

---

### Upgrading

- There is a fix in the syntax for the output of ‘ceph osd tree –format=json’.
- The MDS disk format has changed from prior releases *and* from v0.57. In particular, upgrades to v0.56.4 are safe, but you cannot move from v0.56.4 to v0.57 if you are using the MDS for CephFS; you must upgrade directly to v0.58 (or later) instead.

### Notable changes

- mon: fix bug in bringup with IPv6
- reduce default memory utilization by internal logging (all daemons)
- rgw: fix for bucket removal
- rgw: reopen logs after log rotation
- rgw: fix multipat upload listing
- rgw: don’t copy object when copied onto self
- osd: fix caps parsing for pools with - or \_
- osd: allow pg log trimming when degraded, scrubbing, recovering (reducing memory consumption)
- osd: fix potential deadlock when ‘journal aio = true’
- osd: various fixes for collection creation/removal, rename, temp collections
- osd: various fixes for PG split
- osd: deep-scrub omap key/value data
- osd: fix rare bug in journal replay
- osd: misc fixes for snapshot tracking
- osd: fix leak in recovery reservations on pool deletion

- osd: fix bug in connection management
- osd: fix for op ordering when rebalancing
- ceph-fuse: report file system size with correct units
- mds: get and set directory layout policies via virtual xattrs
- mds: on-disk format revision (see upgrading note above)
- mkcephfs, init-ceph: close potential security issues with predictable filenames

For more detailed information, see [the complete changelog](#).

## v0.56.3 “bobtail”

This release has several bug fixes surrounding OSD stability. Most significantly, an issue with OSDs being unresponsive shortly after startup (and occasionally crashing due to an internal heartbeat check) is resolved. Please upgrade.

## Upgrading

- A bug was fixed in which the OSDMap epoch for PGs without any IO requests was not recorded. If there are pools in the cluster that are completely idle (for example, the `data` and `metadata` pools normally used by CephFS), and a large number of OSDMap epochs have elapsed since the `ceph-osd` daemon was last restarted, those maps will get reprocessed when the daemon restarts. This process can take a while if there are a lot of maps. A workaround is to ‘touch’ any idle pools with IO prior to restarting the daemons after packages are upgraded:

```
1. rados bench 10 write -t 1 -b 4096 -p {POOLNAME}
```

This will typically generate enough IO to touch every PG in the pool without generating significant cluster load, and also cleans up any temporary objects it creates.

## Notable changes

- osd: flush peering work queue prior to start
- osd: persist osdmap epoch for idle PGs
- osd: fix and simplify connection handling for heartbeats
- osd: avoid crash on invalid admin command
- mon: fix rare races with monitor elections and commands

- mon: enforce that OSD reweights be between 0 and 1 (NOTE: not CRUSH weights)
- mon: approximate client, recovery bandwidth logging
- radosgw: fixed some XML formatting to conform to Swift API inconsistency
- radosgw: fix usage accounting bug; add repair tool
- radosgw: make fallback URI configurable (necessary on some web servers)
- librbd: fix handling for interrupted ‘unprotect’ operations
- mds, ceph-fuse: allow file and directory layouts to be modified via virtual xattrs

For more detailed information, see [the complete changelog](#).

## v0.56.2 “bobtail”

---

This release has a wide range of bug fixes, stability improvements, and some performance improvements. Please upgrade.

## Upgrading

---

- The meaning of the ‘osd scrub min interval’ and ‘osd scrub max interval’ has changed slightly. The min interval used to be meaningless, while the max interval would only trigger a scrub if the load was sufficiently low. Now, the min interval option works the way the old max interval did (it will trigger a scrub after this amount of time if the load is low), while the max interval will force a scrub regardless of load. The default options have been adjusted accordingly. If you have customized these in ceph.conf, please review their values when upgrading.
- CRUSH maps that are generated by default when calling `ceph-mon --mkfs` directly now distribute replicas across hosts instead of across OSDs. Any provisioning tools that are being used by Ceph may be affected, although probably for the better, as distributing across hosts is a much more commonly sought behavior. If you use `mkcephfs` to create the cluster, the default CRUSH rule is still inferred by the number of hosts and/or racks in the initial ceph.conf.

## Notable changes

---

- osd: snapshot trimming fixes
- osd: scrub snapshot metadata
- osd: fix osdmap trimming

- osd: misc peering fixes
- osd: stop heartbeating with peers if internal threads are stuck/hung
- osd: PG removal is friendlier to other workloads
- osd: fix recovery start delay (was causing very slow recovery)
- osd: fix scheduling of explicitly requested scrubs
- osd: fix scrub interval config options
- osd: improve recovery vs client io tuning
- osd: improve ‘slow request’ warning detail for better diagnosis
- osd: default CRUSH map now distributes across hosts, not OSDs
- osd: fix crash on 32-bit hosts triggered by librbd clients
- librbd: fix error handling when talking to older OSDs
- mon: fix a few rare crashes
- ceph command: ability to easily adjust CRUSH tunables
- radosgw: object copy does not copy source ACLs
- rados command: fix omap command usage
- sysvinit script: set ulimit -n properly on remote hosts
- msgr: fix narrow race with message queuing
- fixed compilation on some old distros (e.g., RHEL 5.x)

For more detailed information, see [the complete changelog](#).

## v0.56.1 “bobtail”

This release has two critical fixes. Please upgrade.

## Upgrading

- There is a protocol compatibility problem between v0.56 and any other version that is now fixed. If your radosgw or RBD clients are running v0.56, they will need to be upgraded too. If they are running a version prior to v0.56, they can be left as is.

## Notable changes

- osd: fix commit sequence for XFS, ext4 (or any other non-btrfs) to prevent data loss on power cycle or kernel panic
- osd: fix compatibility for CALL operation
- osd: process old osdmaps prior to joining cluster (fixes slow startup)
- osd: fix a couple of recovery-related crashes
- osd: fix large io requests when journal is in (non-default) aio mode
- log: fix possible deadlock in logging code

For more detailed information, see [the complete changelog](#).

## v0.56 “bobtail”

Bobtail is the second stable release of Ceph, named in honor of the Bobtail Squid:  
[https://en.wikipedia.org/wiki/Bobtail\\_squid](https://en.wikipedia.org/wiki/Bobtail_squid).

## Key features since v0.48 “argonaut”

- Object Storage Daemon (OSD): improved threading, small-io performance, and performance during recovery
- Object Storage Daemon (OSD): regular “deep” scrubbing of all stored data to detect latent disk errors
- RADOS Block Device (RBD): support for copy-on-write clones of images.
- RADOS Block Device (RBD): better client-side caching.
- RADOS Block Device (RBD): advisory image locking
- Rados Gateway (RGW): support for efficient usage logging/scraping (for billing purposes)
- Rados Gateway (RGW): expanded S3 and Swift API coverage (e.g., POST, multi-object delete)
- Rados Gateway (RGW): improved striping for large objects
- Rados Gateway (RGW): OpenStack Keystone integration
- RPM packages for Fedora, RHEL/CentOS, OpenSUSE, and SLES
- mkcephfs: support for automatically formatting and mounting XFS and ext4 (in addition to btrfs)

# Upgrading

Please refer to the document [Upgrading from Argonaut to Bobtail](#) for details.

- Cephx authentication is now enabled by default (since v0.55). Upgrading a cluster without adjusting the Ceph configuration will likely prevent the system from starting up on its own. We recommend first modifying the configuration to indicate that authentication is disabled, and only then upgrading to the latest version:

```
1. auth client required = none
2. auth service required = none
3. auth cluster required = none
```

- Ceph daemons can be upgraded one-by-one while the cluster is online and in service.
- The `ceph-osd` daemons must be upgraded and restarted *before* any `radosgw` daemons are restarted, as they depend on some new ceph-osd functionality. (The `ceph-mon`, `ceph-osd`, and `ceph-mds` daemons can be upgraded and restarted in any order.)
- Once each individual daemon has been upgraded and restarted, it cannot be downgraded.
- The cluster of `ceph-mon` daemons will migrate to a new internal on-wire protocol once all daemons in the quorum have been upgraded. Upgrading only a majority of the nodes (e.g., two out of three) may expose the cluster to a situation where a single additional failure may compromise availability (because the non-upgraded daemon cannot participate in the new protocol). We recommend not waiting for an extended period of time between `ceph-mon` upgrades.
- The ops log and usage log for radosgw are now off by default. If you need these logs (e.g., for billing purposes), you must enable them explicitly. For logging of all operations to objects in the `.log` pool (see `radosgw-admin log ...`):

```
1. rgw enable ops log = true
```

For usage logging of aggregated bandwidth usage (see `radosgw-admin usage ...`):

```
1. rgw enable usage log = true
```

- You should not create or use “format 2” RBD images until after all `ceph-osd` daemons have been upgraded. Note that “format 1” is still the default. You can use the new `ceph osd ls` and `ceph tell osd.N version` commands to doublecheck your cluster. `ceph osd ls` will give a list of all OSD IDs that are part of the cluster, and you can use that to write a simple shell loop to display all the OSD version strings:

```

1. for i in $(ceph osd ls); do
2.     ceph tell osd.${i} version
3. done

```

## Compatibility changes

- The ‘ceph osd create [<uuid>]’ command now rejects an argument that is not a UUID. (Previously it would take an optional integer OSD id.) This correct syntax has been ‘ceph osd create [<uuid>]’ since v0.47, but the older calling convention was being silently ignored.
- The CRUSH map root nodes now have type `root` instead of type `pool`. This avoids confusion with RADOS pools, which are not directly related. Any scripts or tools that use the `ceph osd crush ...` commands may need to be adjusted accordingly.
- The `ceph osd pool create <poolname> <pnum>` command now requires the `pnum` argument. Previously this was optional, and would default to 8, which was almost never a good number.
- Degraded mode (when there are fewer than the desired number of replicas) is now more configurable on a per-pool basis, with the `min_size` parameter. By default, with `min_size 0`, this allows I/O to objects with  $N - \text{floor}(N/2)$  replicas, where  $N$  is the total number of expected copies. Argonaut behavior was equivalent to having `min_size = 1`, so I/O would always be possible if any completely up-to-date copy remained. `min_size = 1` could result in lower overall availability in certain cases, such as flapping network partitions.
- The sysvinit start/stop script now defaults to adjusting the max open files `ulimit` to 16384. On most systems the default is 1024, so this is an increase and won’t break anything. If some system has a higher initial value, however, this change will lower the limit. The value can be adjusted explicitly by adding an entry to the `ceph.conf` file in the appropriate section. For example:

```

1. [global]
2.     max open files = 32768

```

- ‘rbd lock list’ and ‘rbd showmapped’ no longer use tabs as separators in their output.
- There is a configurable limit on the number of PGs when creating a new pool, to prevent a user from accidentally specifying a ridiculous number for `pg_num`. It can be adjusted via the ‘mon max pool pg num’ option on the monitor, and defaults to 65536 (the current max supported by the Linux kernel client).
- The osd capabilities associated with a rados user have changed syntax since 0.48 argonaut. The new format is mostly backwards compatible, but there are two backwards-incompatible changes:

- specifying a list of pools in one grant, i.e. ‘allow r pool=foo,bar’ is now done in separate grants, i.e. ‘allow r pool=foo, allow r pool=bar’.
- restricting pool access by pool owner (‘allow r uid=foo’) is removed. This feature was not very useful and unused in practice.

The new format is documented in the ceph-authtool man page.

- ‘rbd cp’ and ‘rbd rename’ use rbd as the default destination pool, regardless of what pool the source image is in. Previously they would default to the same pool as the source image.
- ‘rbd export’ no longer prints a message for each object written. It just reports percent complete like other long-lasting operations.
- ‘ceph osd tree’ now uses 4 decimal places for weight so output is nicer for humans
- Several monitor operations are now idempotent:
  - ceph osd pool create
  - ceph osd pool delete
  - ceph osd pool mksnap
  - ceph osd rm
  - ceph pg <pgid> revert

## Notable changes

---

- auth: enable cephx by default
- auth: expanded authentication settings for greater flexibility
- auth: sign messages when using cephx
- build fixes for Fedora 18, CentOS/RHEL 6
- ceph: new ‘osd ls’ and ‘osd tell <osd.N> version’ commands
- ceph-debugpack: misc improvements
- ceph-disk-prepare: creates and labels GPT partitions
- ceph-disk-prepare: support for external journals, default mount/mkfs options, etc.
- ceph-fuse/libcephfs: many misc fixes, admin socket debugging
- ceph-fuse: fix handling for .. in root directory

- ceph-fuse: many fixes (including memory leaks, hangs)
- ceph-fuse: mount helper (mount.fuse.ceph) for use with /etc/fstab
- ceph.spec: misc packaging fixes
- common: thread pool sizes can now be adjusted at runtime
- config: \$pid is now available as a metavariable
- crush: default root of tree type is now ‘root’ instead of ‘pool’ (to avoid confusiong wrt rados pools)
- crush: fixed retry behavior with chooseleaf via tunable
- crush: tunables documented; feature bit now present and enforced
- libcephfs: java wrapper
- librados: several bug fixes (rare races, locking errors)
- librados: some locking fixes
- librados: watch/notify fixes, misc memory leaks
- librbd: a few fixes to ‘discard’ support
- librbd: fine-grained striping feature
- librbd: fixed memory leaks
- librbd: fully functional and documented image cloning
- librbd: image (advisory) locking
- librbd: improved caching (of object non-existence)
- librbd: ‘flatten’ command to sever clone parent relationship
- librbd: ‘protect’//‘unprotect’ commands to prevent clone parent from being deleted
- librbd: clip requests past end-of-image.
- librbd: fixes an issue with some windows guests running in qemu (remove floating point usage)
- log: fix in-memory buffering behavior (to only write log messages on crash)
- mds: fix into release on abort session close, relative getattr path, mds shutdown, other misc items
- mds: misc fixes
- mkcephfs: fix for default keyring, osd data/journal locations

- mkcephfs: support for formatting xfs, ext4 (as well as btrfs)
- init: support for automatically mounting xfs and ext4 osd data directories
- mon, radosgw, ceph-fuse: fixed memory leaks
- mon: improved ENOSPC, fs error checking
- mon: less-destructive ceph-mon -mkfs behavior
- mon: misc fixes
- mon: more informative info about stuck PGs in ‘health detail’
- mon: information about recovery and backfill in ‘pg <pgid> query’
- mon: new ‘osd crush create-or-move ...’ command
- mon: new ‘osd crush move ...’ command lets you rearrange your CRUSH hierarchy
- mon: optionally dump ‘osd tree’ in json
- mon: configurable cap on maximum osd number (mon max osd)
- mon: many bug fixes (various races causing ceph-mon crashes)
- mon: new on-disk metadata to facilitate future mon changes (post-bobtail)
- mon: election bug fixes
- mon: throttle client messages (limit memory consumption)
- mon: throttle osd flapping based on osd history (limits osdmap ‘thrashing’ on overloaded or unhappy clusters)
- mon: ‘report’ command for dumping detailed cluster status (e.g., for use when reporting bugs)
- mon: osdmap flags like noup, noin now cause a health warning
- msgr: improved failure handling code
- msgr: many bug fixes
- osd, mon: honor new ‘nobackfill’ and ‘norecover’ osdmap flags
- osd, mon: use feature bits to lock out clients lacking CRUSH tunables when they are in use
- osd: backfill reservation framework (to avoid flooding new osds with backfill data)
- osd: backfill target reservations (improve performance during recovery)

- osd: better tracking of recent slow operations
- osd: capability grammar improvements, bug fixes
- osd: client vs recovery io prioritization
- osd: crush performance improvements
- osd: default journal size to 5 GB
- osd: experimental support for PG “splitting” (pg\_num adjustment for existing pools)
- osd: fix memory leak on certain error paths
- osd: fixed detection of EIO errors from fs on read
- osd: major refactor of PG peering and threading
- osd: many bug fixes
- osd: more/better dump info about in-progress operations
- osd: new caps structure (see compatibility notes)
- osd: new ‘deep scrub’ will compare object content across replicas (once per week by default)
- osd: new ‘lock’ rados class for generic object locking
- osd: optional ‘min’ pg size
- osd: recovery reservations
- osd: scrub efficiency improvement
- osd: several out of order reply bug fixes
- osd: several rare peering cases fixed
- osd: some performance improvements related to request queuing
- osd: use entire device if journal is a block device
- osd: use syncfs(2) when kernel supports it, even if glibc does not
- osd: various fixes for out-of-order op replies
- rados: ability to copy, rename pools
- rados: bench command now cleans up after itself
- rados: ‘cppool’ command to copy rados pools

- rados: ‘rm’ now accepts a list of objects to be removed
- radosgw: POST support
- radosgw: REST API for managing usage stats
- radosgw: fix bug in bucket stat updates
- radosgw: fix copy-object vs attributes
- radosgw: fix range header for large objects, ETag quoting, GMT dates, other compatibility fixes
- radosgw: improved garbage collection framework
- radosgw: many small fixes, cleanups
- radosgw: openstack keystone integration
- radosgw: stripe large (non-multipart) objects
- radosgw: support for multi-object deletes
- radosgw: support for swift manifest objects
- radosgw: vanity bucket dns names
- radosgw: various API compatibility fixes
- rbd: import from stdin, export to stdout
- rbd: new ‘ls -l’ option to view images with metadata
- rbd: use generic id and keyring options for ‘rbd map’
- rbd: don’t issue usage on errors
- udev: fix symlink creation for rbd images containing partitions
- upstart: job files for all daemon types (not enabled by default)
- wireshark: ceph protocol dissector patch updated

## v0.54

---

## Upgrading

---

- The osd capabilities associated with a rados user have changed syntax since 0.48 argonaut. The new format is mostly backwards compatible, but there are two backwards-incompatible changes:

- specifying a list of pools in one grant, i.e. ‘allow r pool=foo,bar’ is now done in separate grants, i.e. ‘allow r pool=foo, allow r pool=bar’.
- restricting pool access by pool owner (‘allow r uid=foo’) is removed. This feature was not very useful and unused in practice.

The new format is documented in the `ceph-authtool` man page.

- Bug fixes to the new osd capability format parsing properly validate the allowed operations. If an existing rados user gets permissions errors after upgrading, its capabilities were probably misconfigured. See the `ceph-authtool` man page for details on osd capabilities.
- ‘rbd lock list’ and ‘rbd showmapped’ no longer use tabs as separators in their output.

# v0.48.3 “argonaut”

This release contains a critical fix that can prevent data loss or corruption after a power loss or kernel panic event. Please upgrade immediately.

## Upgrading

- If you are using the undocumented `ceph-disk-prepare` and `ceph-disk-activate` tools, they have several new features and some additional functionality. Please review the changes in behavior carefully before upgrading.
- The .deb packages now require `xfsprogs`.

## Notable changes

- filestore: fix op\_seq write order (fixes journal replay after power loss)
- osd: fix occasional indefinitely hung “slow” request
- osd: fix encoding for `pool_snap_info_t` when talking to pre-v0.48 clients
- osd: fix heartbeat check
- osd: reduce log noise about rbd watch
- log: fixes for deadlocks in the internal logging code
- log: make log buffer size adjustable
- init script: fix for ‘ceph status’ across machines
- radosgw: fix swift error handling
- radosgw: fix swift authentication concurrency bug
- radosgw: don’t cache large objects
- radosgw: fix some memory leaks
- radosgw: fix timezone conversion on read
- radosgw: relax date format restrictions
- radosgw: fix multipart overwrite
- radosgw: stop processing requests on client disconnect
- radosgw: avoid adding port to url that already has a port

- radosgw: fix copy to not override ETAG
- common: make parsing of ip address lists more forgiving
- common: fix admin socket compatibility with old protocol (for collectd plugin)
- mon: drop dup commands on paxos reset
- mds: fix loner selection for multiclient workloads
- mds: fix compat bit checks
- ceph-fuse: fix segfault on startup when keyring is missing
- ceph-authtool: fix usage
- ceph-disk-activate: misc backports
- ceph-disk-prepare: misc backports
- debian: depend on xfsprogs (we use xfs by default)
- rpm: build rpms, some related Makefile changes

For more detailed information, see [the complete changelog](#).

## v0.48.2 “argonaut”

---

### Upgrading

---

- The default search path for keyring files now includes /etc/ceph/ceph.\$name.keyring. If such files are present on your cluster, be aware that by default they may now be used.
- There are several changes to the upstart init files. These have not been previously documented or recommended. Any existing users should review the changes before upgrading.
- The ceph-disk-prepare and ceph-disk-active scripts have been updated significantly. These have not been previously documented or recommended. Any existing users should review the changes before upgrading.

### Notable changes

---

- mkcephfs: fix keyring generation for mds, osd when default paths are used
- radosgw: fix bug causing occasional corruption of per-bucket stats
- radosgw: workaround to avoid previously corrupted stats from going negative

- radosgw: fix bug in usage stats reporting on busy buckets
- radosgw: fix Content-Range: header for objects bigger than 2 GB.
- rbd: avoid leaving watch acting when command line tool errors out (avoids 30s delay on subsequent operations)
- rbd: friendlier use of --pool/-image options for import (old calling convention still works)
- librbd: fix rare snapshot creation race (could “lose” a snap when creation is concurrent)
- librbd: fix discard handling when spanning holes
- librbd: fix memory leak on discard when caching is enabled
- objecter: misc fixes for op reordering
- objecter: fix for rare startup-time deadlock waiting for osdmap
- ceph: fix usage
- mon: reduce log noise about “check\_sub”
- ceph-disk-activate: misc fixes, improvements
- ceph-disk-prepare: partition and format osd disks automatically
- upstart: start everyone on a reboot
- upstart: always update the osd crush location on start if specified in the config
- config: add /etc/ceph/ceph.\$name.keyring to default keyring search path
- ceph.spec: don’t package crush headers

For more detailed information, see [the complete changelog](#).

## v0.48.1 “argonaut”

---

### Upgrading

- The radosgw usage trim function was effectively broken in v0.48. Earlier it would remove more usage data than what was requested. This is fixed in v0.48.1, but the fix is incompatible. The v0.48 radosgw-admin tool cannot be used to initiate the trimming; please use the v0.48.1 version.
- v0.48.1 now explicitly indicates support for the CRUSH\_TUNABLES feature. No other version of Ceph requires this, yet, but future versions will when the tunables

are adjusted from their historical defaults.

- There are no other compatibility changes between v0.48.1 and v0.48.

## Notable changes

---

- mkcephfs: use default ‘keyring’, ‘osd data’, ‘osd journal’ paths when not specified in conf
- msgr: various fixes to socket error handling
- osd: reduce scrub overhead
- osd: misc peering fixes (past\_interval sharing, pgs stuck in ‘peering’ states)
- osd: fail on EIO in read path (do not silently ignore read errors from failing disks)
- osd: avoid internal heartbeat errors by breaking some large transactions into pieces
- osd: fix osdmap catch-up during startup (catch up and then add daemon to osdmap)
- osd: fix spurious ‘misdirected op’ messages
- osd: report scrub status via ‘pg ... query’
- rbd: fix race when watch registrations are resent
- rbd: fix rbd image id assignment scheme (new image data objects have slightly different names)
- rbd: fix perf stats for cache hit rate
- rbd tool: fix off-by-one in key name (crash when empty key specified)
- rbd: more robust udev rules
- rados tool: copy object, pool commands
- radosgw: fix in usage stats trimming
- radosgw: misc API compatibility fixes (date strings, ETag quoting, swift headers, etc.)
- ceph-fuse: fix locking in read/write paths
- mon: fix rare race corrupting on-disk data
- config: fix admin socket ‘config set’ command
- log: fix in-memory log event gathering

- debian: remove crush headers, include librados-config
- rpm: add ceph-disk-{activate, prepare}

For more detailed information, see [the complete changelog](#).

## v0.48 “argonaut”

### Upgrading

- This release includes a disk format upgrade. Each ceph-osd daemon, upon startup, will migrate its locally stored data to the new format. This process can take a while (for large object counts, even hours), especially on non-btrfs file systems.
- To keep the cluster available while the upgrade is in progress, we recommend you upgrade a storage node or rack at a time, and wait for the cluster to recover each time. To prevent the cluster from moving data around in response to the OSD daemons being down for minutes or hours, you may want to:

```
1. ceph osd set noout
```

This will prevent the cluster from marking down OSDs as “out” and re-replicating the data elsewhere. If you do this, be sure to clear the flag when the upgrade is complete:

```
1. ceph osd unset noout
```

- There is a encoding format change internal to the monitor cluster. The monitor daemons are careful to switch to the new format only when all members of the quorum support it. However, that means that a partial quorum with new code may move to the new format, and a recovering monitor running old code will be unable to join (it will crash). If this occurs, simply upgrading the remaining monitor will resolve the problem.
- The ceph tool’s -s and -w commands from previous versions are incompatible with this version. Upgrade your client tools at the same time you upgrade the monitors if you rely on those commands.
- It is not possible to downgrade from v0.48 to a previous version.

### Notable changes

- osd: stability improvements

- osd: capability model simplification
- osd: simpler/safer -mkfs (no longer removes all files; safe to re-run on active osd)
- osd: potentially buggy FIEMAP behavior disabled by default
- rbd: caching improvements
- rbd: improved instrumentation
- rbd: bug fixes
- radosgw: new, scalable usage logging infrastructure
- radosgw: per-user bucket limits
- mon: streamlined process for setting up authentication keys
- mon: stability improvements
- mon: log message throttling
- doc: improved documentation (ceph, rbd, radosgw, chef, etc.)
- config: new default locations for daemon keyrings
- config: arbitrary variable substitutions
- improved ‘admin socket’ daemon admin interface (ceph -admin-daemon ...)
- chef: support for multiple monitor clusters
- upstart: basic support for monitors, mds, radosgw; osd support still a work in progress.

The new default keyring locations mean that when enabling authentication (`auth supported = cephx`), keyring locations do not need to be specified if the keyring file is located inside the daemon’s data directory (`/var/lib/ceph/$type/ceph-$id` by default).

There is also a lot of librbd code in this release that is laying the groundwork for the upcoming layering functionality, but is not actually used. Likewise, the upstart support is still incomplete and not recommended; we will backport that functionality later if it turns out to be non-disruptive.

# Ceph Glossary

---

Ceph is growing rapidly. As firms deploy Ceph, the technical terms such as "RADOS", "RBD," "RGW" and so forth require corresponding marketing terms that explain what each component does. The terms in this glossary are intended to complement the existing technical terminology.

Sometimes more than one term applies to a definition. Generally, the first term reflects a term consistent with Ceph's marketing, and secondary terms reflect either technical terms or legacy ways of referring to Ceph systems.

## Ceph Project

The aggregate term for the people, software, mission and infrastructure of Ceph.

## cephx

The Ceph authentication protocol. Cephx operates like Kerberos, but it has no single point of failure.

## Ceph

## Ceph Platform

All Ceph software, which includes any piece of code hosted at <https://github.com/ceph>.

## Ceph System

## Ceph Stack

A collection of two or more components of Ceph.

## Ceph Node

## Node

## Host

Any single machine or server in a Ceph System.

## Ceph Storage Cluster

## Ceph Object Store

## RADOS

## RADOS Cluster

## Reliable Autonomic Distributed Object Store

The core set of storage software which stores the user's data (MON+OSD).

## Ceph Cluster Map

### Cluster Map

The set of maps comprising the monitor map, OSD map, PG map, MDS map and CRUSH map. See [Cluster Map](#) for details.

## Ceph Object Storage

The object storage “product”, service or capabilities, which consists essentially of a Ceph Storage Cluster and a Ceph Object Gateway.

## Ceph Object Gateway

### RADOS Gateway

### RGW

The S3/Swift gateway component of Ceph.

## Ceph Block Device

### RBD

The block storage component of Ceph.

## Ceph Block Storage

The block storage “product,” service or capabilities when used in conjunction with `librbd`, a hypervisor such as QEMU or Xen, and a hypervisor abstraction layer such as `libvirt`.

## Ceph File System

### CephFS

### Ceph FS

The POSIX filesystem components of Ceph. Refer [CephFS Architecture](#) and [Ceph File System](#) for more details.

## Cloud Platforms

### Cloud Stacks

Third party cloud provisioning platforms such as OpenStack, CloudStack, OpenNebula, ProxMox, etc.

## Object Storage Device

### OSD

A physical or logical storage unit (e.g., LUN). Sometimes, Ceph users use the term

“OSD” to refer to [Ceph OSD Daemon](#), though the proper term is “Ceph OSD”.

Ceph OSD Daemon

Ceph OSD Daemons

Ceph OSD

The Ceph OSD software, which interacts with a logical disk ([OSD](#)). Sometimes, Ceph users use the term “OSD” to refer to “Ceph OSD Daemon”, though the proper term is “Ceph OSD”.

OSD id

The integer that defines an OSD. It is generated by the monitors as part of the creation of a new OSD.

OSD fsid

This is a unique identifier used to further improve the uniqueness of an OSD and it is found in the OSD path in a file called [osd\\_fsid](#). This [fsid](#) term is used interchangeably with [uuid](#)

OSD uuid

Just like the OSD fsid, this is the OSD unique identifier and is used interchangeably with [fsid](#)

bluestore

OSD BlueStore is a new back end for OSD daemons (kraken and newer versions). Unlike [filestore](#) it stores objects directly on the Ceph block devices without any file system interface.

filestore

A back end for OSD daemons, where a Journal is needed and files are written to the filesystem.

Ceph Monitor

MON

The Ceph monitor software.

Ceph Manager

MGR

The Ceph manager software, which collects all the state from the whole cluster in one place.

Ceph Manager Dashboard

Ceph Dashboard

Dashboard Module

Dashboard Plugin

Dashboard

A built-in web-based Ceph management and monitoring application to administer various aspects and objects of the cluster. The dashboard is implemented as a Ceph Manager module. See [Ceph Dashboard](#) for more details.

Ceph Metadata Server

MDS

The Ceph metadata software.

Ceph Clients

Ceph Client

The collection of Ceph components which can access a Ceph Storage Cluster. These include the Ceph Object Gateway, the Ceph Block Device, the Ceph File System, and their corresponding libraries, kernel modules, and FUSEs.

Ceph Kernel Modules

The collection of kernel modules which can be used to interact with the Ceph System (e.g., `ceph.ko` , `rbd.ko` ).

Ceph Client Libraries

The collection of libraries that can be used to interact with components of the Ceph System.

Ceph Release

Any distinct numbered version of Ceph.

Ceph Point Release

Any ad-hoc release that includes only bug or security fixes.

Ceph Interim Release

Versions of Ceph that have not yet been put through quality assurance testing, but may contain new features.

Ceph Release Candidate

A major version of Ceph that has undergone initial quality assurance testing and is ready for beta testers.

#### Ceph Stable Release

A major version of Ceph where all features from the preceding interim releases have been put through quality assurance testing successfully.

#### Ceph Test Framework

#### Teuthology

The collection of software that performs scripted tests on Ceph.

#### CRUSH

Controlled Replication Under Scalable Hashing. It is the algorithm Ceph uses to compute object storage locations.

#### CRUSH rule

The CRUSH data placement rule that applies to a particular pool(s).

#### Pool

#### Pools

Pools are logical partitions for storing objects.

#### systemd oneshot

A systemd `type` where a command is defined in `ExecStart` which will exit upon completion (it is not intended to daemonize)

#### LVM tags

Extensible metadata for LVM volumes and groups. It is used to store Ceph-specific information about devices and its relationship with OSDs.