UPGRADING CEPH

Each release of Ceph may have additional steps. Refer to the release-specific sections in this document and the release notes document to identify release-specific procedures for your cluster before using the upgrade procedures.

SUMMARY

You can upgrade daemons in your Ceph cluster while the cluster is online and in service! Certain types of daemons depend upon others. For example, Ceph Metadata Servers and Ceph Object Gateways depend upon Ceph Monitors and Ceph OSD Daemons. We recommend upgrading in this order:

1. Ceph Deploy
2. Ceph Monitors
3. Ceph OSD Daemons
4. Ceph Metadata Servers
5. Ceph Object Gateways

As a general rule, we recommend upgrading all the daemons of a specific type (e.g., all `ceph-mon` daemons, all `ceph-osd` daemons, etc.) to ensure that they are all on the same release. We also recommend that you upgrade all the daemons in your cluster before you try to exercise new functionality in a release.

The Upgrade Procedures are relatively simple, but please look at distribution-specific sections before upgrading. The basic process involves three steps:

1. Use `ceph-deploy` on your admin node to upgrade the packages for multiple hosts (using the `ceph-deploy install` command), or login to each host and upgrade the Ceph package manually. For example, when Upgrading Monitors, the `ceph-deploy` syntax might look like this:

   ```
   ceph-deploy install --release {release-name} ceph-node1[ ceph-node2]
   ceph-deploy install --release firefly mon1 mon2 mon3
   ```

   Note: The `ceph-deploy install` command will upgrade the packages in the specified node(s) from the old release to the release you specify. There is no `ceph-deploy upgrade` command.

2. Login in to each Ceph node and restart each Ceph daemon. See Operating a Cluster for details.

3. Ensure your cluster is healthy. See Monitoring a Cluster for details.

Important: Once you upgrade a daemon, you cannot downgrade it.

CEPH DEPLOY

Before upgrading Ceph daemons, upgrade the `ceph-deploy` tool.

```
sudo pip install -U ceph-deploy
```

Or:

```
sudo apt-get install ceph-deploy
```

Or:

```
sudo yum install ceph-deploy python-pushy
```

ARGONAUT TO BOBTAIL

When upgrading from Argonaut to Bobtail, you need to be aware of several things:

1. Authentication now defaults to **ON**, but used to default to **OFF**.
2. Monitors use a new internal on-wire protocol.
3. RBD `format2` images require upgrading all OSDs before using it.

Ensure that you update package repository paths. For example:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-bobtail/ $(lsb_release -sc) main | sudo tee /etc/apt
```

See the following sections for additional details.

AUTHENTICATION

The Ceph Bobtail release enables authentication by default. Bobtail also has finer-grained authentication configuration settings. In previous versions of Ceph (i.e., actually v 0.55 and earlier), you could simply specify:

```
auth supported = [cephx | none]
```

This option still works, but is deprecated. New releases support `cluster`, `service` and `client` authentication settings as follows:

```
auth cluster required = [cephx | none]  # default cephx
auth service required = [cephx | none] # default cephx
auth client required = [cephx | none] # default cephx,none
```

Important: If your cluster does not currently have an `auth supported` line that enables authentication, you must explicitly turn it off in Bobtail using the settings below.:

```
auth cluster required = none
auth service required = none
```

This will disable authentication on the cluster, but still leave clients with the default configuration where they can talk to a cluster that does enable it, but do not require it.

Important: If your cluster already has an `auth supported` option defined in the configuration file, no changes are necessary.

See User Management - Backward Compatibility for details.

MONITOR ON-WIRE PROTOCOL

We recommend upgrading all monitors to Bobtail. A mixture of Bobtail and Argonaut monitors will not be able to use the new on-wire protocol, as the protocol requires all monitors to be Bobtail or greater. 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.

RBD IMAGES

The Bobtail release supports `format 2` images! However, 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:

```
for i in $(ceph osd ls); do
    ceph tell osd.${i} version
done
```

ARGONAUT TO CUTTLEFISH

To upgrade your cluster from Argonaut to Cuttlefish, please read this section, and the sections on upgrading from Argonaut to Bobtail and upgrading from Bobtail to Cuttlefish carefully. When upgrading from Argonaut to Cuttlefish, **YOU MUST UPGRADE YOUR MONITORS FROM ARGONAUT TO BOBTAIL v0.56.5 FIRST!!!**. All other Ceph daemons can upgrade from Argonaut to Cuttlefish without the intermediate upgrade to Bobtail.

Important: Ensure that the repository specified points to Bobtail, not Cuttlefish.

For example:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-bobtail/ $(lsb_release -sc) main | sudo tee /etc/apt
```

We recommend upgrading all monitors to Bobtail before proceeding with the upgrade of the monitors to Cuttlefish. A mixture of Bobtail and Argonaut monitors will not be able to use the new on-wire protocol, as the protocol requires all monitors to be Bobtail or greater. 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. See Upgrading Monitors for details.

Note: See the Authentication section and the User Management - Backward Compatibility for additional information on authentication backward compatibility settings for Bobtail.

Once you complete the upgrade of your monitors from Argonaut to Bobtail, and have restarted the monitor daemons, you must upgrade the monitors from Bobtail to Cuttlefish. Ensure that you have a quorum before beginning this upgrade procedure. Before upgrading, remember to replace the reference to the Bobtail repository with a reference to the Cuttlefish repository. For example:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-cuttlefish/ $(lsb_release -sc) main | sudo tee /etc/
```

See Upgrading Monitors for details.

The architecture of the monitors changed significantly from Argonaut to Cuttlefish. See Monitor Config Reference and Joao's blog post for details. Once you complete the monitor upgrade, you can upgrade the OSD daemons and the MDS daemons using the generic procedures. See Upgrading an OSD and Upgrading a Metadata Server for details.

BOBTAIL TO CUTTLEFISH

Upgrading your cluster from Bobtail to Cuttlefish has a few important considerations. First, the monitor uses a new architecture, so you should upgrade the full set of monitors to use Cuttlefish. Second, if you run multiple metadata servers in a cluster, ensure the metadata servers have unique names. See the following sections for details.

Replace any `apt` reference to older repositories with a reference to the Cuttlefish repository. For example:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-cuttlefish/ $(lsb_release -sc) main | sudo tee /etc/
```

MONITOR

The architecture of the monitors changed significantly from Bobtail to Cuttlefish. See Monitor Config Reference and Joao's blog post for details. This means that v0.59 and pre-v0.59 monitors do not talk to each other (Cuttlefish is v.0.61). When you upgrade each monitor, it will convert its local data store to the new format. Once you upgrade a majority of monitors, the monitors form a quorum using the new protocol and the old monitors will be blocked until they get upgraded. For this reason, we recommend upgrading the monitors in immediate succession.

Important: Do not run a mixed-version cluster for an extended period.

MDS UNIQUE NAMES

The monitor now enforces that MDS names be unique. If you have multiple metadata server daemons that start with the same ID (e.g., mds.a) the second metadata server will implicitly mark the first metadata server as `failed`. Multi-MDS configurations with identical names must be adjusted accordingly to give daemons unique names. If you run your cluster with one metadata server, you can disregard this notice for now.

CEPH-DEPLOY

The `ceph-deploy` tool is now the preferred method of provisioning new clusters. For existing clusters created via the obsolete `mkcephfs` tool that would like to transition to the new tool, there is a migration path, documented at Transitioning to ceph-deploy.

CUTTLEFISH TO DUMPLING

When upgrading from Cuttlefish (v0.61-v0.61.7) you may perform a rolling upgrade. However, there are a few important considerations. First, you must upgrade the ceph command line utility, because it has changed significantly. Second, you must upgrade the full set of monitors to use Dumpling, because of a protocol change.

Replace any reference to older repositories with a reference to the Dumpling repository. For example, with `apt` perform the following:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-dumpling/ $(lsb_release -sc) main | sudo tee /etc/ap
```

With CentOS/Red Hat distributions, remove the old repository.

```
sudo rm /etc/yum.repos.d/ceph.repo
```

Then add a new `ceph.repo` repository entry with the following contents.

```
[ceph]
name=Ceph Packages and Backports $basearch
baseurl=http://download.ceph.com/rpm/el6/$basearch
enabled=1
gpgcheck=1
gpgkey=https://download.ceph.com/keys/release.asc
```

Note: Ensure you use the correct URL for your distribution. Check the http://download.ceph.com/rpm directory for your distribution.

Note: Since you can upgrade using `ceph-deploy` you will only need to add the repository on Ceph Client nodes where you use the ceph command line interface or the `ceph-deploy` tool.

DUMPLING TO EMPEROR

When upgrading from Dumpling (v0.64) you may perform a rolling upgrade.

Replace any reference to older repositories with a reference to the Emperor repository. For example, with `apt` perform the following:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-emperor/ $(lsb_release -sc) main | sudo tee /etc/apt
```

With CentOS/Red Hat distributions, remove the old repository.

```
sudo rm /etc/yum.repos.d/ceph.repo
```

Then add a new `ceph.repo` repository entry with the following contents and replace `{distro}` with your distribution (e.g., `el6`, `rhel6`, etc).

```
[ceph]
name=Ceph Packages and Backports $basearch
baseurl=http://download.ceph.com/rpm-emperor/{distro}/$basearch
enabled=1
gpgcheck=1
gpgkey=https://download.ceph.com/keys/release.asc
```

Note: Ensure you use the correct URL for your distribution. Check the http://download.ceph.com/rpm directory for your distribution.

Note: Since you can upgrade using `ceph-deploy` you will only need to add the repository on Ceph Client nodes where you use the ceph command line interface or the `ceph-deploy` tool.

COMMAND LINE UTILITY

In V0.65, the ceph commandline interface (CLI) utility changed significantly. You will not be able to use the old CLI with Dumpling. This means that you must upgrade the `ceph-common` library on all nodes that access the Ceph Storage Cluster with the ceph CLI before upgrading Ceph daemons.

```
sudo apt-get update && sudo apt-get install ceph-common
```

Ensure that you have the latest version (v0.67 or later). If you do not, you may need to uninstall, auto remove dependencies and reinstall.

See v0.65 for details on the new command line interface.

MONITOR

Dumpling (v0.67) `ceph-mon` daemons have an internal protocol change. This means that v0.67 daemons cannot talk to v0.66 or older daemons. Once you upgrade a majority of monitors, the monitors form a quorum using the new protocol and the old monitors will be blocked until they get upgraded. For this reason, we recommend upgrading all monitors at once (or in relatively quick succession) to minimize the possibility of downtime.

Important: Do not run a mixed-version cluster for an extended period.

DUMPLING TO FIREFLY

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.

MONITOR

Dumpling (v0.67) `ceph-mon` daemons have an internal protocol change. This means that v0.67 daemons cannot talk to v0.66 or older daemons. Once you upgrade a majority of monitors, the monitors form a quorum using the new protocol and the old monitors will be blocked until they get upgraded. For this reason, we recommend upgrading all monitors at once (or in relatively quick succession) to minimize the possibility of downtime.

Important: Do not run a mixed-version cluster for an extended period.

CEPH CONFIG FILE CHANGES

We recommand adding the following to the `[mon]` section of your `ceph.conf` prior to upgrade:

```
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, 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.

COMMAND LINE UTILITY

In V0.65, the ceph commandline interface (CLI) utility changed significantly. You will not be able to use the old CLI with Firefly. This means that you must upgrade the `ceph-common` library on all nodes that access the Ceph Storage Cluster with the ceph CLI before upgrading Ceph daemons.

For Debian/Ubuntu, execute:

```
sudo apt-get update && sudo apt-get install ceph-common
```

For CentOS/RHEL, execute:

```
sudo yum install ceph-common
```

Ensure that you have the latest version. If you do not, you may need to uninstall, auto remove dependencies and reinstall.

See v0.65 for details on the new command line interface.

UPGRADE SEQUENCE

Replace any reference to older repositories with a reference to the Firely repository. For example, with `apt` perform the following:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-firefly/ $(lsb_release -sc) main | sudo tee /etc/apt
```

With CentOS/Red Hat distributions, remove the old repository.

```
sudo rm /etc/yum.repos.d/ceph.repo
```

Then add a new `ceph.repo` repository entry with the following contents and replace `{distro}` with your distribution (e.g., `el6`, `rhel6`, `rhel7`, etc.).

```
[ceph]
name=Ceph Packages and Backports $basearch
baseurl=http://download.ceph.com/rpm-firefly/{distro}/$basearch
enabled=1
gpgcheck=1
gpgkey=https://download.ceph.com/keys/release.asc
```

Upgrade daemons in the following order:

1. **Monitors:** If the `ceph-mon` daemons are not restarted prior to the `ceph-osd` daemons, the monitors will not correctly register their new capabilities with the cluster and new features may not be usable until the monitors are restarted a second time.
2. **OSDs**
3. **MDSs:** If the `ceph-mds` daemon is restarted first, it will wait until all OSDs have been upgraded before finishing its startup

sequence.

4. **Gateways:** 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`.

Note: Make sure you upgrade your **ALL** of your Ceph monitors **AND** restart them **BEFORE** upgrading and restarting OSDs, MDSs, and gateways!

EMPEROR TO FIREFLY

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 Cuttlefish to Dumpling and the Firefly release notes for details. To upgrade from a post-Emperor point release, see the Firefly release notes for details.

CEPH CONFIG FILE CHANGES

We recommand adding the following to the `[mon]` section of your `ceph.conf` prior to upgrade:

```
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, 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 SEQUENCE

Replace any reference to older repositories with a reference to the Firefly repository. For example, with `apt` perform the following:

```
sudo rm /etc/apt/sources.list.d/ceph.list
echo deb http://download.ceph.com/debian-firefly/ $(lsb_release -sc) main | sudo tee /etc/apt
```

With CentOS/Red Hat distributions, remove the old repository.

```
sudo rm /etc/yum.repos.d/ceph.repo
```

Then add a new `ceph.repo` repository entry with the following contents, but replace `{distro}` with your distribution (e.g., `el6`, `rhel6`, `rhel7`, etc.).

```
[ceph]
name=Ceph Packages and Backports $basearch
baseurl=http://download.ceph.com/rpm/{distro}/$basearch
enabled=1
gpgcheck=1
gpgkey=https://download.ceph.com/keys/release.asc
```

Note: Ensure you use the correct URL for your distribution. Check the http://download.ceph.com/rpm directory for your distribution.

Note: Since you can upgrade using `ceph-deploy` you will only need to add the repository on Ceph Client nodes where you use the `ceph` command line interface or the `ceph-deploy` tool.

Upgrade daemons in the following order:

1. **Monitors:** If the `ceph-mon` daemons are not restarted prior to the `ceph-osd` daemons, the monitors will not correctly register their new capabilities with the cluster and new features may not be usable until the monitors are restarted a second time.
2. **OSDs**

3. **MDSs:** If the `ceph-mds` daemon is restarted first, it will wait until all OSDs have been upgraded before finishing its startup sequence.
4. **Gateways:** 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`.

UPGRADE PROCEDURES

The following sections describe the upgrade process.

Important: Each release of Ceph may have some additional steps. Refer to release-specific sections for details **BEFORE** you begin upgrading daemons.

UPGRADING MONITORS

To upgrade monitors, perform the following steps:

1. Upgrade the Ceph package for each daemon instance.

   You may use `ceph-deploy` to address all monitor nodes at once. For example:

   ```
   ceph-deploy install --release {release-name} ceph-node1[ ceph-node2]
   ceph-deploy install --release hammer mon1 mon2 mon3
   ```

   You may also use the package manager for your Linux distribution on each individual node. To upgrade packages manually on each Debian/Ubuntu host, perform the following steps .

   ```
   ssh {mon-host}
   sudo apt-get update && sudo apt-get install ceph
   ```

   On CentOS/Red Hat hosts, perform the following steps:

   ```
   ssh {mon-host}
   sudo yum update && sudo yum install ceph
   ```

2. Restart each monitor. For Ubuntu distributions, use:

   ```
   sudo restart ceph-mon id={hostname}
   ```

   For CentOS/Red Hat/Debian distributions, use:

   ```
   sudo /etc/init.d/ceph restart {mon-id}
   ```

   For CentOS/Red Hat distributions deployed with `ceph-deploy`, the monitor ID is usually `mon.{hostname}`.

3. Ensure each monitor has rejoined the quorum.

   ```
   ceph mon stat
   ```

Ensure that you have completed the upgrade cycle for all of your Ceph Monitors.

UPGRADING AN OSD

To upgrade a Ceph OSD Daemon, perform the following steps:

1. Upgrade the Ceph OSD Daemon package.

   You may use `ceph-deploy` to address all Ceph OSD Daemon nodes at once. For example:

```
ceph-deploy install --release {release-name} ceph-node1[ ceph-node2]
ceph-deploy install --release hammer osd1 osd2 osd3
```

You may also use the package manager on each node to upgrade packages manually. For Debian/Ubuntu hosts, perform the following steps on each host.

```
ssh {osd-host}
sudo apt-get update && sudo apt-get install ceph
```

For CentOS/Red Hat hosts, perform the following steps:

```
ssh {osd-host}
sudo yum update && sudo yum install ceph
```

2. Restart the OSD, where `N` is the OSD number. For Ubuntu, use:

```
sudo restart ceph-osd id=N
```

For multiple OSDs on a host, you may restart all of them with Upstart.

```
sudo restart ceph-osd-all
```

For CentOS/Red Hat/Debian distributions, use:

```
sudo /etc/init.d/ceph restart N
```

3. Ensure each upgraded Ceph OSD Daemon has rejoined the cluster:

```
ceph osd stat
```

Ensure that you have completed the upgrade cycle for all of your Ceph OSD Daemons.

UPGRADING A METADATA SERVER

To upgrade a Ceph Metadata Server, perform the following steps:

1. Upgrade the Ceph Metadata Server package. You may use `ceph-deploy` to address all Ceph Metadata Server nodes at once, or use the package manager on each node. For example:

```
ceph-deploy install --release {release-name} ceph-node1
ceph-deploy install --release hammer mds1
```

To upgrade packages manually, perform the following steps on each Debian/Ubuntu host.

```
ssh {mon-host}
sudo apt-get update && sudo apt-get install ceph-mds
```

Or the following steps on CentOS/Red Hat hosts:

```
ssh {mon-host}
sudo yum update && sudo yum install ceph-mds
```

2. Restart the metadata server. For Ubuntu, use:

```
sudo restart ceph-mds id={hostname}
```

For CentOS/Red Hat/Debian distributions, use:

```
sudo /etc/init.d/ceph restart mds.{hostname}
```

For clusters deployed with ceph-deploy, the name is usually either the name you specified on creation or the hostname.

3. Ensure the metadata server is up and running:

```
ceph mds stat
```

UPGRADING A CLIENT

Once you have upgraded the packages and restarted daemons on your Ceph cluster, we recommend upgrading ceph-common and client libraries (librbd1 and librados2) on your client nodes too.

1. Upgrade the package:

```
ssh {client-host}
apt-get update && sudo apt-get install ceph-common librados2 librbd1 python-rados python-
```

2. Ensure that you have the latest version:

```
ceph --version
```

If you do not have the latest version, you may need to uninstall, auto remove dependencies and reinstall.

TRANSITIONING TO CEPH-DEPLOY

If you have an existing cluster that you deployed with mkcephfs (usually Argonaut or Bobtail releases), you will need to make a few changes to your configuration to ensure that your cluster will work with ceph-deploy.

MONITOR KEYRING

You will need to add caps mon = "allow *" to your monitor keyring if it is not already in the keyring. By default, the monitor keyring is located under /var/lib/ceph/mon/ceph-$id/keyring. When you have added the caps setting, your monitor keyring should look something like this:

```
[mon.]
        key = AQBJIHhRuHCwDRAAZjBTSJcIBIoGpdOR9ToiyQ==
        caps mon = "allow *"
```

Adding caps mon = "allow *" will ease the transition from mkcephfs to ceph-deploy by allowing ceph-create-keys to use the mon. keyring file in $mon_data and get the caps it needs.

USE DEFAULT PATHS

Under the /var/lib/ceph directory, the mon and osd directories need to use the default paths.

- **OSDs**: The path should be /var/lib/ceph/osd/ceph-$id
- **MON**: The path should be /var/lib/ceph/mon/ceph-$id

Under those directories, the keyring should be in a file named keyring.