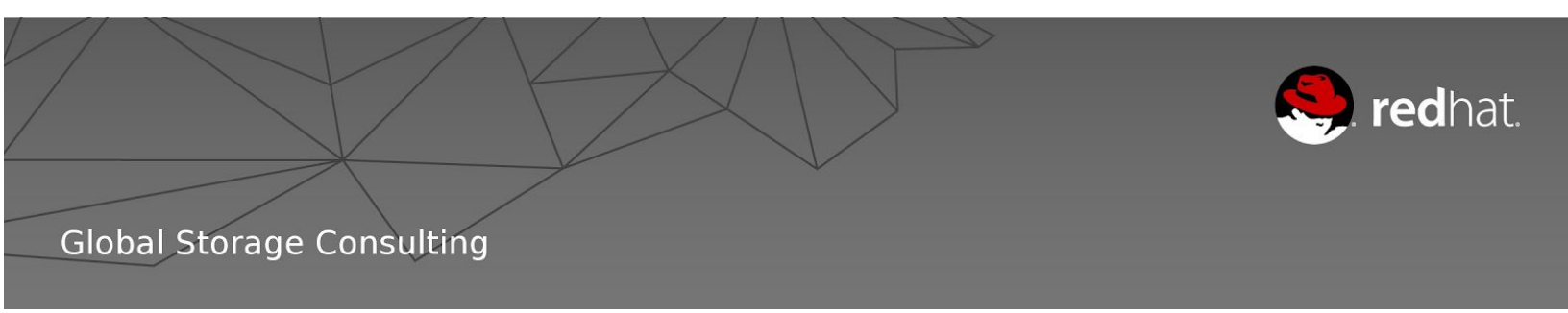

Red Hat Ceph Storage 2.3 Installation Guide

PREPARED FOR

Wichita State University

Prepared on 2017-06-30 by:

Alexandre Marangone	Senior Storage Consultant	amarango@redhat.com

Document v1.0

1. Preface

1.1 Confidentiality, Copyright, and Disclaimer

This is a Customer-facing document between Red Hat, Inc. and Wichita State University ("Client").

1.2 About This Document

This document provides the reader with a detailed description of the results of the Red Hat Ceph Storage Installation performed by Red Hat Consulting for Wichita State University.

1.3 Terms

Term	Definition
OSD	Object Storage Daemon
CRUSH	Controlled Replication Under Scalable Hashing
MON	Monitor Node
RADOS	Reliable, Autonomous, Distributed Object Storage
RGW	RADOS Gateway
RHCS	Red Hat Ceph Storage
RHSC	Red Hat Storage Console

Table 1.4-1: Terminology

2. Introduction

This document describes the steps taken in order to install Red Hat Ceph Storage 2.0 at WSU.

WSU intends to use the OpenStack Cinder, Glance and Nova services to present storage to tenants who require storage within their OpenStack environments. WSU also intends to provide Object Storage capabilities to OpenStack instances which require it by using the Ceph Object Gateway (RGW) and the Swift API it exposes.

3. Next steps and consideration

Issue during this engagement prevented us from doing the following:

- Register the Ceph nodes: No access to a NAT'ed VLAN.
- Move the Ceph Monitors to the OSP Controllers: The OSP Overcloud might have to be redeployed to reflect a network change.
- Adding a PCI-e NVMe on each physical box and using it as a Ceph journal would improve small random write performance.

3. Cluster Installation

3.1 Convention

In this document, every command line will look like `<HOST># <CMD>` where host represent the hostname of the machine, or a group of machines. For instance `ALL` means that this command need to be run on all hosts. `ALL_OSD` means that the command need to be run on all storage nodes.

3.2 Prerequisite

Setup NTP:

```
ALL# yum install -y ntp
 #configure /etc/ntp.conf with local NTP server(s) (ntp1.wichita.edu,
ntp2.wichita.edu)
ALL# ntpdate <ntp_server>
ALL# systemctl start ntpd
ALL# systemctl enable ntpd
```

From ceph1, create a SSH key and add it to the other hosts:

```
ceph1# ssh-keygen
ceph1# ssh-copy-id ceph2
ceph1# ssh-copy-id ceph3
ceph1# ssh-copy-id os-radosgw
```

NOTE Make sure entries are present in `/etc/hosts`

Install the ceph-ansible package on ceph1:

```
ceph1# yum install -y ceph-ansible
```

NOTE We used a local clone of the `rhel-7-server-rhscon-2-installer-rpms`. After registering the nodes, WSU will need to enable this repo on this node.

Configure the Ansible inventory file on ceph1:

```
ceph1# cat /etc/ansible/hosts
[mons]
ceph1
ceph2
ceph3

[osds]
ceph1
ceph2
ceph3
```

```
[rgws]
os-radosgw
```

Configure firewall from ceph1:

```
ceph1# ansible mons -a "firewall-cmd --zone=public --add-port=6789/tcp"
ceph1# ansible mons -a "firewall-cmd --zone=public --add-port=6789/tcp
--permanent"
ceph1# ansible osds -a "firewall-cmd --zone=public --add-port=6900-7300/tcp"
ceph1# ansible osds -a "firewall-cmd --zone=public --add-port=6900-7300/tcp
--permanent"
ceph1# ansible rgws -a "firewall-cmd --zone=public --add-port=80/tcp"
ceph1# ansible rgws -a "firewall-cmd --zone=public --add-port=80/tcp
--permanent"
```

3.3 Cluster deployment

We will deploy this cluster using ceph-ansible from the ceph1 node.

Create a directory for Ansible to store Ceph keys:

```
ceph1# mkdir /root/ceph-ansible-keys
```

Create ceph-ansible variable files

```
ceph1# cd /usr/share/ceph-ansible
ceph1# cp group_vars/all.yml.sample group_vars/all.yml
ceph1# cp group_vars/osds.sample group_vars/osds.yml
ceph1# cp group_vars/rgws.sample group_vars/rgws.yml
ceph1# cp site.yml.sample site.yml
```

Configure variables. In this output we will only show the modification we made and not the full files:

```
ceph1# egrep -v "^#|^$" /usr/share/ceph-ansible/group_vars/all
---
dummy:
fetch_directory: /root/ceph-ansible-keys
ceph_rhcs: true
ceph_rhcs_version: 2
ceph_rhcs_iso_install: true
ceph_rhcs_iso_path: /root/ceph_stuff/rhceph-2.3-rhel-7-x86_64.iso
ceph_rhcs_mount_path: /mnt/rhcs2
ceph_rhcs_repository_path: "{{ ceph_stable_rh_storage_repository_path |
default('/tmp/rh-storage-repo') }}" # where to copy iso's content
monitor_interface: bond0.3254
ip_version: ipv4
journal_size: 5120 # OSD journal size in MB
public_network: 192.168.254.0/24
```

```
cluster_network: 192.168.255.0/24
osd_mount_options_xfs:
noatime,largeio,inode64,swalloc,context="system_u:object_r:ceph_var_lib_t:s0"
radosgw_civetweb_port: 80
radosgw_civetweb_bind_ip: 0.0.0.0 # when using ipv6 enclose with brackets:
"[{{ ansible_default_ipv6.address }}]"
radosgw_civetweb_num_threads: 200
ceph_conf_overrides:
    client.rgw.os-radosgw:
        rgw_keystone_url: http://192.168.251.50:35357
        rgw_keystone_admin_token: GyQc49GsxvaCDMWH9NwU4Gc4J
        rgw_keystone_accepted_roles: Member, admin, _member_
        rgw_s3_auth_use_keystone: true
        #nss_db_path: /var/ceph/nss
        rgw_keystone_verify_ssl: false
        rgw_dns_name: objects.cs.wichita.edu
        rgw_ops_log_rados: false
```

```
ceph1# egrep -v "^#|^$" /usr/share/ceph-ansible/group_vars/osds
---
dummy:
devices:
  - /dev/sda
  - /dev/sdb
  - /dev/sdc
  - /dev/sdd
  - /dev/sde
  - /dev/sdf
  - /dev/sdg
  - /dev/sdh
  - /dev/sdi
  - /dev/sdj
  - /dev/sdk
  - /dev/sdl
journal_collocation: true
```

In the above configuration, we tell ceph-ansible about our drives. The devices variable defines all the drives that will be used to store data. We will collocate the Ceph journal by creating a small partition (5GB) on each of them.

```
ceph1# egrep -v '^#|^$' group_vars/rgws.yml
---
dummy:
copy_admin_key: true
rgw_override_bucket_index_max_shards: 16
```

In the above configuration, we copy the admin keyring to the radosgw host for ease of use.

Deploy the cluster by running the Ansible playbook:

```
ceph1# cd /usr/share/ceph-ansible
ceph1# ansible-playbook site.yml
```

3.4 Configuring Ceph for integration with OpenStack

Set the CRUSH tunables to optimal, this ensure the best possible data distribution:

```
ANY_MON# ceph osd crush tunables optimal
```

Note that non RHCS 2.0 clients (i.e. prior Ceph versions) will not be able to communicate with the Ceph cluster with tunables set to optimal.

Create the pools for the RGW (Swift/S3) and OpenStack block:

```
ceph1# ceph osd pool create volumes 1024
ceph1# ceph osd pool create images 128
ceph1# ceph osd pool create vms 256
ceph1# ceph osd pool create backups 128
ceph1# ceph osd pool create default.rgw.buckets.data 128
ceph1# ceph osd pool create default.rgw.buckets.index 64
ceph1# ceph osd pool create metrics 64
```

NOTE The Object Gateway will create other pools. The default amount of PG each pool will be created with (8) is sufficient for a cluster this size.

Create a Cephx key for OpenStack integration

```
ANY_MON# ceph auth get-or-create client.openstack mon 'allow r' osd 'allow
class-read object_prefix rbd_children, allow rwx pool=volumes, allow rwx
pool=vms, allow rwx pool=images, allow rwx pool=backups, allow rwx
pool=metrics'
```

The key generated will be used in the OSP-d configuration template for OpenStack integration.

3.5 Using the Object Gateway with Openstack

All the configuration required on the Ceph side was done during the radosgw deployment. See the all.yml file that specifies rgw_keystone_* configuration options.
The last step required is to update OSP's endpoints for the Object-Store service.

Check if endpoints already exist:

```
heat-admin@undercloud$ source ~/overcloudrc.v3
heat-admin@undercloud$ openstack endpoint list
```

If endpoints exist delete them one by one with `openstack endpoint delete <service_id>`.

Add the new endpoints:

```
heat-admin@undercloud$ source ~/overcloudrc.v3
heat-admin@undercloud$ openstack endpoint create --region regionOne swift
public http://192.168.251.14:80/swift/v1
heat-admin@undercloud$ openstack endpoint create --region regionOne swift
internal http://192.168.251.14:80/swift/v1
heat-admin@undercloud$ openstack endpoint create --region regionOne swift
admin http://192.168.251.14:80/swift/v1
```

NOTE 192.168.251.14 is the IP of the radosgw

4. Recreating RadosGW endpoints

If the overcloud needs to be redeployed at some point, the RadosGW endpoint will need to be updated.

The following settings will need to be updated on `ceph1`, in `/usr/share/ceph-ansible/group_vars/all.yml`:

- `rgw_keystone_url: http://192.168.251.50:35357`
- `rgw_keystone_admin_token: GyQc49GsxvaCDMWH9NwU4Gc4J`

Once updated, run the playbook with:

```
ceph1# cd /usr/share/ceph-ansible
ceph1# ansible-playbook site.yml
```

Finally, recreate the endpoint as described in Section 3.5 “Using the Object Gateway with Openstack”.

5. Appendix II - Common operations

Red Hat recommends using the Red Hat Ceph Storage administration guide to reference all operations on the cluster. The Administration guide will remain up to date for a specific version of RHCS. We have included a quick reference for some common operations below.

https://access.redhat.com/documentation/en/red-hat-ceph-storage/1.3/administration-guide/administration-guide

- Check cluster health

```
[mon_node]# ceph -s
```

  HEALTH_OK shows everything is running as it should, HEALTH_WARN shows that something is happening to the cluster but IO are still possible. HEALTH_ERR requires immediate attention as IO might be blocked. Other details will be visible

- When a HEALTH_WARN or HEALTH_ERR is displayed, this command will show more detail about the error

```
[mon_node]# ceph health detail
```

- Display the CRUSH hierarchy of the cluster as long as informations on each OSD (up/down, in/out).

```
[mon_node]# ceph osd tree
```

- Viewing pools (multiple methods ordered by usefulness)

```
[mon_node]# ceph osd dump | grep pool
```

```
[mon_node]# ceph df
```

```
[mon_node]# ceph osd pool ls
```

We recommend the following reading on Ceph

Ceph Architecture
http://docs.ceph.com/docs/master/architecture/
Ceph Object Gateway documentation
http://docs.ceph.com/docs/master/radosgw/
Add/Remove Monitors
http://docs.ceph.com/docs/master/rados/operations/add-or-rm-mons/
Ceph command reference
http://docs.ceph.com/docs/master/rados/operations/control/
RHCS 2.0 configuration guide:
https://access.redhat.com/documentation/en/red-hat-ceph-storage/2/paged/configuration-guide/
RHCS 2.0 administration guide:
https://access.redhat.com/documentation/en/red-hat-ceph-storage/2/paged/administration-guide
Ceph Troubleshooting
http://docs.ceph.com/docs/master/rados/troubleshooting/