Red Hat Ceph Storage 7

Block Device to OpenStack Guide

Configuring Ceph, QEMU, libvirt and OpenStack to use Ceph as a back end for
OpenStack.

Last Updated: 2024-03-08

Red Hat Ceph Storage 7 Block Device to OpenStack Guide

Configuring Ceph, QEMU, libvirt and OpenStack to use Ceph as a back end for OpenStack.

Legal Notice

Abstract

This document describes how to configure OpenStack and Ceph to use Ceph as a back end for Glance, Cinder, Cinder Backup and Nova. Red Hat is committed to replacing problematic language in our code, documentation, and web properties. We are beginning with these four terms: master, slave, blacklist, and whitelist. Because of the enormity of this endeavor, these changes will be implemented gradually over several upcoming releases. For more details, see our CTO Chris Wright's message

Table of Contents

CHAPTER 1. CEPH BLOCK DEVICES AND OPENSTACK

The Red Hat Enterprise Linux OpenStack Platform Director provides two methods for using Ceph as a backend for Glance, Cinder, Cinder Backup and Nova:

1. **OpenStack creates the Ceph storage cluster:** OpenStack Director can create a Ceph storage cluster. This requires configuring templates for the Ceph OSDs. OpenStack handles the installation and configuration of Ceph hosts. With this scenario, OpenStack will install the Ceph monitors with the OpenStack controller hosts.

2. **OpenStack connects to an existing Ceph storage cluster:** OpenStack Director, using Red Hat OpenStack Platform 9 and higher, can connect to a Ceph monitor and configure the Ceph storage cluster for use as a backend for OpenStack.

The foregoing methods are the preferred methods for configuring Ceph as a backend for OpenStack, because they will handle much of the installation and configuration automatically.

This document details the manual procedure for configuring Ceph, QEMU, libvirt and OpenStack to use Ceph as a backend. This document is intended for use for those who do not intend to use the RHEL OSP Director.
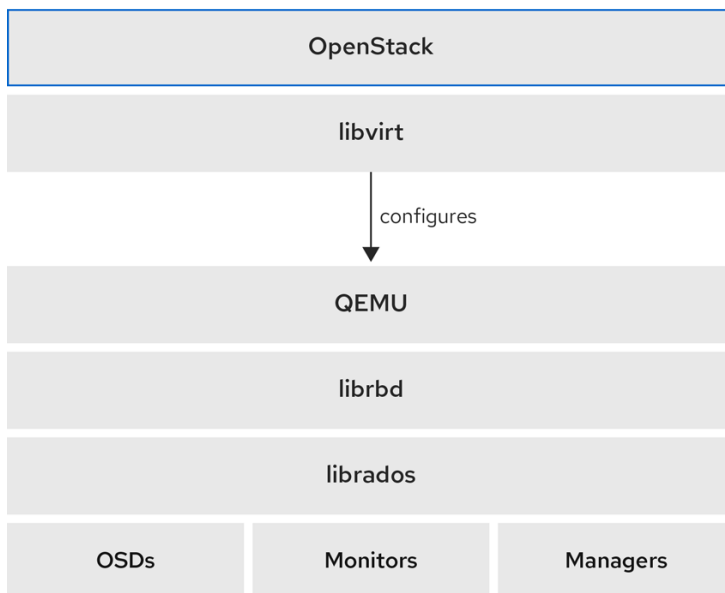
NOTE

A running Ceph storage cluster and at least one OpenStack host is required to use Ceph block devices as a backend for OpenStack.

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. Ceph can serve as a black end for OpenStack Cinder and Cinder Backup.

- **Guest Disks:** Guest disks are guest operating system disks. By default, when booting a virtual

machine, its disk appears as a file on the file system of the hypervisor, by default, under **/var/lib/nova/instances/<uuid>/** directory. OpenStack Glance can store images in a Ceph block device, and can use Cinder to boot a virtual machine using a copy-on-write clone of an image.
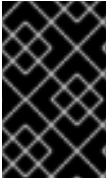
> **IMPORTANT**
>
> Ceph doesn't support QCOW2 for hosting a virtual machine disk. To boot virtual machines, either ephemeral backend or booting from a volume, the Glance image format must be RAW.

OpenStack can use Ceph for images, volumes or guest disks virtual machines. There is no requirement for using all three.

Additional Resources

- See the Red Hat OpenStack Platform documentation for additional details.

CHAPTER 2. INSTALLING AND CONFIGURING CEPH FOR OPENSTACK

As a storage administrator, you must install and configure Ceph before the Red Hat OpenStack Platform can use the Ceph block devices.

Prerequisites

- A new or existing Red Hat Ceph Storage cluster.

2.1. CREATING CEPH POOLS FOR OPENSTACK

You can create Ceph pools for use with OpenStack. By default, Ceph block devices use the **rbd** pool, but you can use any available pool.

Prerequisites

- A running Red Hat Ceph Storage cluster.

Procedure

1. Verify the Red Hat Ceph Storage cluster is running, and is in a **HEALTH_OK** state:

   ```
   [root@mon ~]# ceph -s
   ```

2. Create the Ceph pools:

   Example

   ```
   [root@mon ~]# ceph osd pool create volumes 128
   [root@mon ~]# ceph osd pool create backups 128
   [root@mon ~]# ceph osd pool create images 128
   [root@mon ~]# ceph osd pool create vms 128
   ```

   In the above example, **128** is the number of placement groups.

   > **IMPORTANT**
   >
   > Red Hat recommends using the Ceph Placement Group's per Pool Calculator to calculate a suitable number of placement groups for the pools.

Additional Resources

- See the Pools chapter in the *Storage Strategies* guide for more details on creating pools.

2.2. INSTALLING THE CEPH CLIENT ON OPENSTACK

You can install the Ceph client packages on the Red Hat OpenStack Platform to access the Ceph storage cluster.

Prerequisites

- A running Red Hat Ceph Storage cluster.

- Access to the Ceph software repository.

- Root-level access to the OpenStack Nova, Cinder, Cinder Backup and Glance nodes.

Procedure

1. On the OpenStack Nova, Cinder, Cinder Backup nodes install the following packages:

   ```
   [root@nova ~]# dnf install python-rbd ceph-common
   ```

2. On the OpenStack Glance host install the **python-rbd** package:

   ```
   [root@glance ~]# dnf install python-rbd
   ```

2.3. COPYING THE CEPH CONFIGURATION FILE TO OPENSTACK

Copying the Ceph configuration file to the **nova-compute**, **cinder-backup**, **cinder-volume**, and **glance-api** nodes.

Prerequisites

- A running Red Hat Ceph Storage cluster.

- Access to the Ceph software repository.

- Root-level access to the OpenStack Nova, Cinder, and Glance nodes.

Procedure

1. Copy the Ceph configuration file from the Ceph Monitor host to the OpenStack Nova, Cinder, Cinder Backup and Glance nodes:

   ```
   [root@mon ~]# scp /etc/ceph/ceph.conf OPENSTACK_NODES:/etc/ceph
   ```

2.4. CONFIGURING CEPH CLIENT AUTHENTICATION

You can configure authentication for the Ceph client to access the Red Hat OpenStack Platform.

Prerequisites

- Root-level access to the Ceph Monitor host.

- A running Red Hat Ceph Storage cluster.

Procedure

1. From a Ceph Monitor host, create new users for Cinder, Cinder Backup and Glance:

   ```
   [root@mon ~]# ceph auth get-or-create client.cinder mon 'allow r' osd 'allow class-read
   object_prefix rbd_children, allow rwx pool=volumes, allow rwx pool=vms, allow rx
   pool=images'
   ```

```
[root@mon ~]# ceph auth get-or-create client.cinder-backup mon 'allow r' osd 'allow class-read object_prefix rbd_children, allow rwx pool=backups'

[root@mon ~]# ceph auth get-or-create client.glance mon 'allow r' osd 'allow class-read object_prefix rbd_children, allow rwx pool=images'
```

2. Add the keyrings for **client.cinder**, **client.cinder-backup** and **client.glance** to the appropriate nodes and change their ownership:

```
[root@mon ~]# ceph auth get-or-create client.cinder | ssh CINDER_VOLUME_NODE sudo tee /etc/ceph/ceph.client.cinder.keyring
[root@mon ~]# ssh CINDER_VOLUME_NODE chown cinder:cinder /etc/ceph/ceph.client.cinder.keyring

[root@mon ~]# ceph auth get-or-create client.cinder-backup | ssh CINDER_BACKUP_NODE tee /etc/ceph/ceph.client.cinder-backup.keyring
[root@mon ~]# ssh CINDER_BACKUP_NODE chown cinder:cinder /etc/ceph/ceph.client.cinder-backup.keyring

[root@mon ~]# ceph auth get-or-create client.glance | ssh GLANCE_API_NODE sudo tee /etc/ceph/ceph.client.glance.keyring
[root@mon ~]# ssh GLANCE_API_NODE chown glance:glance /etc/ceph/ceph.client.glance.keyring
```

3. OpenStack Nova nodes need the keyring file for the **nova-compute** process:

```
[root@mon ~]# ceph auth get-or-create client.cinder | ssh NOVA_NODE tee /etc/ceph/ceph.client.cinder.keyring
```

4. The OpenStack Nova nodes also need to store the secret key of the **client.cinder** user in **libvirt**. The **libvirt** process needs the secret key to access the cluster while attaching a block device from Cinder. Create a temporary copy of the secret key on the OpenStack Nova nodes:

```
[root@mon ~]# ceph auth get-key client.cinder | ssh NOVA_NODE tee client.cinder.key
```

If the storage cluster contains Ceph block device images that use the **exclusive-lock** feature, ensure that all Ceph block device users have permissions to blocklist clients:
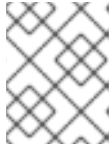
```
[root@mon ~]# ceph auth caps client.ID mon 'allow r, allow command "osd blacklist"' osd 'EXISTING_OSD_USER_CAPS'
```

5. Return to the OpenStack Nova host:

```
[root@mon ~]# ssh NOVA_NODE
```

6. Generate a UUID for the secret, and save the UUID of the secret for configuring **nova-compute** later:

```
[root@nova ~]# uuidgen > uuid-secret.txt
```

> **NOTE**
>
> You do not necessarily need the UUID on all the Nova compute nodes. However, from a platform consistency perspective, it's better to keep the same UUID.

7. On the OpenStack Nova nodes, add the secret key to **libvirt** and remove the temporary copy of the key:

```
cat > secret.xml <<EOF
<secret ephemeral='no' private='no'>
  <uuid>`cat uuid-secret.txt`</uuid>
  <usage type='ceph'>
    <name>client.cinder secret</name>
  </usage>
</secret>
EOF
```

8. Set and define the secret for **libvirt**:

```
[root@nova ~]# virsh secret-define --file secret.xml
[root@nova ~]# virsh secret-set-value --secret $(cat uuid-secret.txt) --base64 $(cat client.cinder.key) && rm client.cinder.key secret.xml
```

Additional Resources

- See the *Managing Ceph users* section in the *Red Hat Ceph Storage Administration Guide* for more details.

- See the *Configuring the existing ceph storage cluster* section in the *Integrating an Overcloud with an Existing Red Hat Ceph Cluster Guide* for Red Hat OpenStack Platform, to know more about user capabilities.

CHAPTER 3. CONFIGURING OPENSTACK TO USE CEPH BLOCK DEVICES

As a storage administrator, you must configure the Red Hat OpenStack Platform to use the Ceph block devices. The Red Hat OpenStack Platform can use Ceph block devices for Cinder, Cinder Backup, Glance, and Nova.

Prerequisites

- A new or existing Red Hat Ceph Storage cluster.

- A running Red Hat OpenStack Platform environment.

3.1. CONFIGURING CINDER TO USE CEPH BLOCK DEVICES

The Red Hat OpenStack Platform can use Ceph block devices to provide back-end storage for Cinder volumes.

Prerequisites

- Root-level access to the Cinder node.

- A Ceph **volume** pool.

- The user and UUID of the secret to interact with Ceph block devices.

Procedure

1. Edit the Cinder configuration file:

   ```
   [root@cinder ~]# vim /etc/cinder/cinder.conf
   ```

2. In the **[DEFAULT]** section, enable Ceph as a backend for Cinder:

   ```
   enabled_backends = ceph
   ```

3. Ensure that the Glance API version is set to 2. If you are configuring multiple cinder back ends in **enabled_backends**, the **glance_api_version = 2** setting must be in the **[DEFAULT]** section and not the **[ceph]** section.

   ```
   glance_api_version = 2
   ```

4. Create a **[ceph]** section in the **cinder.conf** file. Add the Ceph settings in the following steps under the **[ceph]** section.

5. Specify the **volume_driver** setting and set it to use the Ceph block device driver:

   ```
   volume_driver = cinder.volume.drivers.rbd.RBDDriver
   ```

6. Specify the cluster name and Ceph configuration file location. In typical deployments the Ceph cluster has a cluster name of **ceph** and a Ceph configuration file at **/etc/ceph/ceph.conf**. If the Ceph cluster name is not **ceph**, specify the cluster name and configuration file path appropriately:

```
rbd_cluster_name = us-west
rbd_ceph_conf = /etc/ceph/us-west.conf
```

7. By default, Red Hat OpenStack Platform stores Ceph volumes in the **rbd** pool. To use the **volumes** pool created earlier, specify the **rbd_pool** setting and set the **volumes** pool:

```
rbd_pool = volumes
```

8. Red Hat OpenStack Platform does not have a default user name or a UUID of the secret for volumes. Specify **rbd_user** and set it to the **cinder** user. Then, specify the **rbd_secret_uuid** setting and set it to the generated UUID stored in the **uuid-secret.txt** file:

```
rbd_user = cinder
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
```

9. Specify the following settings:

```
rbd_flatten_volume_from_snapshot = false
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1
```

When you configure Cinder to use Ceph block devices, the configuration file might look similar to this:

Example

```
[DEFAULT]
enabled_backends = ceph
glance_api_version = 2
…

[ceph]
volume_driver = cinder.volume.drivers.rbd.RBDDriver
rbd_cluster_name = ceph
rbd_pool = volumes
rbd_user = cinder
rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_flatten_volume_from_snapshot = false
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1
```
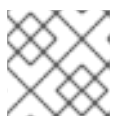
> **NOTE**
>
> Consider removing the default **[lvm]** section and its settings.

3.2. CONFIGURING CINDER BACKUP TO USE CEPH BLOCK DEVICES

The Red Hat OpenStack Platform can configure Cinder backup to use Ceph block devices.

Prerequisites

- Root-level access to the Cinder node.

Procedure

1. Edit the Cinder configuration file:

   ```
   [root@cinder ~]# vim /etc/cinder/cinder.conf
   ```

2. Go to the **[ceph]** section of the configuration file.

3. Specify the **backup_driver** setting and set it to the Ceph driver:

   ```
   backup_driver = cinder.backup.drivers.ceph
   ```

4. Specify the **backup_ceph_conf** setting and specify the path to the Ceph configuration file:

   ```
   backup_ceph_conf = /etc/ceph/ceph.conf
   ```

   > **NOTE**
   >
   > The Cinder backup Ceph configuration file may be different from the Ceph configuration file used for Cinder. For example, it can point to a different Ceph storage cluster.

5. Specify the Ceph pool for backups:

   ```
   backup_ceph_pool = backups
   ```

   > **NOTE**
   >
   > The Ceph configuration file used for Cinder backup might be different from the Ceph configuration file used for Cinder.

6. Specify the **backup_ceph_user** setting and specify the user as **cinder-backup**:

   ```
   backup_ceph_user = cinder-backup
   ```

7. Specify the following settings:

   ```
   backup_ceph_chunk_size = 134217728
   backup_ceph_stripe_unit = 0
   backup_ceph_stripe_count = 0
   restore_discard_excess_bytes = true
   ```

   When you include the Cinder options, the **[ceph]** section of the **cinder.conf** file might look similar to this:

   Example

   ```
   [ceph]
   volume_driver = cinder.volume.drivers.rbd.RBDDriver
   ```

```
rbd_cluster_name = ceph
rbd_pool = volumes
rbd_user = cinder
rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_flatten_volume_from_snapshot = false
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
rbd_max_clone_depth = 5
rbd_store_chunk_size = 4
rados_connect_timeout = -1

backup_driver = cinder.backup.drivers.ceph
backup_ceph_user = cinder-backup
backup_ceph_conf = /etc/ceph/ceph.conf
backup_ceph_chunk_size = 134217728
backup_ceph_pool = backups
backup_ceph_stripe_unit = 0
backup_ceph_stripe_count = 0
restore_discard_excess_bytes = true
```

8. Verify if Cinder backup is enabled:

```
[root@cinder ~]# cat /etc/openstack-dashboard/local_settings | grep enable_backup
```

If **enable_backup** is set to **False**, then edit the **local_settings** file and set it to **True**.

Example

```
OPENSTACK_CINDER_FEATURES = {
    'enable_backup': True,
}
```

3.3. CONFIGURING GLANCE TO USE CEPH BLOCK DEVICES

The Red Hat OpenStack Platform can configure Glance to use Ceph block devices.

Prerequisites

- Root-level access to the Glance node.

Procedure

1. To use Ceph block devices by default, edit the **/etc/glance/glance-api.conf** file. If you used different pool, user or Ceph configuration file settings apply the appropriate values. Uncomment the following settings if necessary and change their values accordingly:

```
[root@glance ~]# vim /etc/glance/glance-api.conf
```

```
stores = rbd
default_store = rbd
rbd_store_chunk_size = 8
rbd_store_pool = images
rbd_store_user = glance
rbd_store_ceph_conf = /etc/ceph/ceph.conf
```

2. To enable copy-on-write (CoW) cloning set **show_image_direct_url** to **True**.

   ```
   show_image_direct_url = True
   ```

   > **IMPORTANT**
   >
   > Enabling CoW exposes the back end location via Glance's API, so the endpoint should not be publicly accessible.

3. Disable cache management if necessary. The **flavor** should be set to **keystone** only, not **keystone+cachemanagement**.

   ```
   flavor = keystone
   ```

4. Red Hat recommends the following properties for images:

   ```
   hw_scsi_model=virtio-scsi
   hw_disk_bus=scsi
   hw_qemu_guest_agent=yes
   os_require_quiesce=yes
   ```

   The **virtio-scsi** controller gets better performance and provides support for discard operations. For systems using SCSI/SAS drives, connect every Cinder block device to that controller. Also, enable the QEMU guest agent and send **fs-freeze/thaw** calls through the QEMU guest agent.

3.4. CONFIGURING NOVA TO USE CEPH BLOCK DEVICES

The Red Hat OpenStack Platform can configure Nova to use Ceph block devices.

You must configure each Nova node to use ephemeral back-end storage devices, which allows all virtual machines to use the Ceph block devices.

Prerequisites

- Root-level access to the Nova nodes.

Procedure

1. Edit the Ceph configuration file:

   ```
   [root@nova ~]# vim /etc/ceph/ceph.conf
   ```

2. Add the following section to the **[client]** section of the Ceph configuration file:

   ```
   [client]
   rbd cache = true
   rbd cache writethrough until flush = true
   rbd concurrent management ops = 20
   admin socket = /var/run/ceph/guests/$cluster-$type.$id.$pid.$cctid.asok
   log file = /var/log/ceph/qemu-guest-$pid.log
   ```

3. Create new directories for the admin socket and log file, and change the directory permissions to use the **qemu** user and **libvirtd** group:

```
[root@nova ~]# mkdir -p /var/run/ceph/guests/ /var/log/ceph/
[root@nova ~]# chown qemu:libvirt /var/run/ceph/guests /var/log/ceph/
```
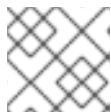
> **NOTE**
>
> The directories must be allowed by SELinux or AppArmor.

4. On each Nova node, edit the **/etc/nova/nova.conf** file. Under the **[libvirt]** section, configure the following settings:

Example

```
[libvirt]
images_type = rbd
images_rbd_pool = vms
images_rbd_ceph_conf = /etc/ceph/ceph.conf
rbd_user = cinder
rbd_secret_uuid = 4b5fd580-360c-4f8c-abb5-c83bb9a3f964
disk_cachemodes="network=writeback"
inject_password = false
inject_key = false
inject_partition = -2
live_migration_flag="VIR_MIGRATE_UNDEFINE_SOURCE,VIR_MIGRATE_PEER2PEER,VIR_MIGRATE_LIVE,VIR_MIGRATE_PERSIST_DEST,VIR_MIGRATE_TUNNELLED"
hw_disk_discard = unmap
```

Replace the UUID in **rbd_user_secret** with the UUID in the **uuid-secret.txt** file.

3.5. RESTARTING THE OPENSTACK SERVICES

Restarting the Red Hat OpenStack Platform services enables you to activate the Ceph block device drivers.

Prerequisites

- Root-level access to the Red Hat OpenStack Platform nodes.

Procedure

1. Load the block device pool names and Ceph user names into the configuration file.

2. Restart the appropriate OpenStack services after modifying the corresponding configuration files:

```
[root@osp ~]# systemctl restart openstack-cinder-volume
[root@osp ~]# systemctl restart openstack-cinder-backup
[root@osp ~]# systemctl restart openstack-glance-api
[root@osp ~]# systemctl restart openstack-nova-compute
```