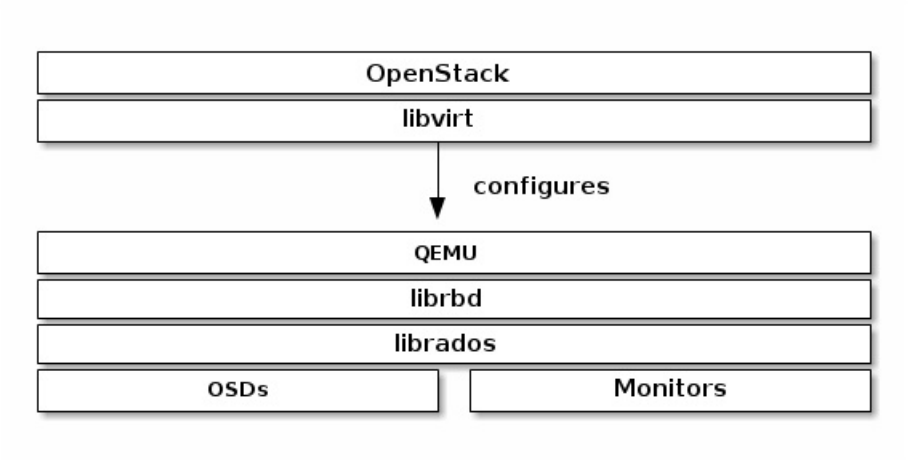
BLOCK DEVICES AND OPENSTACK

You may use Ceph block device images with OpenStack 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 OpenStack, you must install QEMU, `libvirt`, and OpenStack first. We recommend using a separate physical host 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 a running Ceph cluster.

Two 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 `nova-volume` prior to the Folsom release. OpenStack manages volumes using Cinder services beginning with the Folsom release.

Beginning with OpenStack Folsom and Ceph 0.52, you can use OpenStack Glance to store images in a Ceph block device, and you can use Cinder or `nova-volume` to boot a VM using a copy-on-write clone of an image.

The instructions below detail the setup for Glance and Nova/Cinder, 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.

CREATE A POOL

By default, Ceph block devices use the `rbd` pool. You may use any available pool. We recommend creating a pool for Nova/Cinder and a pool for Glance. Ensure your Ceph cluster is running, then create the pools.

```
ceph osd pool create volumes 128
ceph osd pool create images 128
```

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.

CONFIGURE OPENSTACK CEPH CLIENTS

The hosts running `glance-api`, `nova-compute`, and `nova-volume` or `cinder-volume` act as Ceph clients. Each requires the `ceph.conf` file:

```
ssh {your-openstack-server} sudo tee /etc/ceph/ceph.conf </etc/ceph/ceph.conf
```

INSTALL CEPH CLIENT PACKAGES

On the `glance-api` host, you'll need the Python bindings for `librbd`:

```
sudo apt-get install python-ceph
```

On the `nova-volume` or `cinder-volume` host, use the client command line tools:

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

SETUP CEPH CLIENT AUTHENTICATION

If you have cephx authentication enabled, create a new user for Nova/Cinder and Glance.

For Ceph version 0.53 or lower, execute the following:

```
ceph auth get-or-create client.volumes mon 'allow r' osd 'allow x, allow rwx pool=volumes, al
ceph auth get-or-create client.images mon 'allow r' osd 'allow x, allow rwx pool=images'
```

In Ceph version 0.54, more specific permissions were added, so the users can be restricted further. For Ceph version 0.54 or later, execute the following:

```
ceph auth get-or-create client.volumes mon 'allow r' osd 'allow class-read object_prefix rbd_
ceph auth get-or-create client.images mon 'allow r' osd 'allow class-read object_prefix rbd_c
```

Add the keyrings for `client.volumes` and `client.images` to the appropriate hosts and change their ownership:

```
ceph auth get-or-create client.images | ssh {your-glance-api-server} sudo tee /etc/ceph/ceph.
ssh {your-glance-api-server} sudo chown glance:glance /etc/ceph/ceph.client.images.keyring
ceph auth get-or-create client.volumes | ssh {your-volume-server} sudo tee /etc/ceph/ceph.cli
ssh {your-volume-server} sudo chown cinder:cinder /etc/ceph/ceph.client.volumes.keyring
```

Hosts running `nova-compute` do not need the keyring. Instead, they store the secret key in libvirt. Create a temporary copy of the secret key on the hosts running `nova-compute`:

```
ssh {your-compute-host} client.volumes.key <`ceph auth get-key client.volumes`
```

Then, on the compute hosts, add the secret key to libvirt and remove the temporary copy of the key:

```
cat > secret.xml <<EOF
<secret ephemeral='no' private='no'>
  <usage type='ceph'>
    <name>client.volumes secret</name>
  </usage>
</secret>
EOF
sudo virsh secret-define --file secret.xml
<uuid of secret is output here>
sudo virsh secret-set-value --secret {uuid of secret} --base64 $(cat client.volumes.key) && r
```

Save the uuid of the secret for configuring `nova-compute` later.

CONFIGURE OPENSTACK TO USE CEPH

CONFIGURING GLANCE

Glance can use multiple back ends to store images. To use Ceph block devices by default, edit `/etc/glance/glance-api.conf` and add:

```
default_store=rbd
rbd_store_user=images
rbd_store_pool=images
```

If you're using Folsom and want to enable copy-on-write cloning of images into volumes, also add:

```
show_image_direct_url=True
```

Note that this exposes the back end location via Glance's API, so the endpoint with this option enabled should not be publicly accessible.

CONFIGURING CINDER/NOVA-VOLUME

OpenStack requires a driver to interact with Ceph block devices. You must also specify the pool name for the block device. On your OpenStack host, edit `/etc/cinder/cinder.conf` and add:

```
volume_driver=cinder.volume.driver.RBDDriver
rbd_pool=volumes
```

If you're not using Cinder, replace Cinder with Nova in the previous section.

If you're using cephx authentication, also configure the user and uuid of the secret you added to libvirt earlier:

```
rbd_user=volumes
rbd_secret_uuid={uuid of secret}
```

Finally, on each host running `cinder-volume` or `nova-volume`, add `CEPH_ARGS="--id volumes"` to the init/upstart script that starts it.

For example, on Ubuntu, add `env CEPH_ARGS="--id volumes"` to the top of `/etc/init/cinder-volume.conf`.

RESTART OPENSTACK

To activate the Ceph block device driver and load the block device pool name into the configuration, you must restart OpenStack. Navigate the directory where you installed OpenStack, and execute the following:

```
./rejoin-stack.sh
```

If you have OpenStack configured as a service, you can also execute these commands on the appropriate hosts:

```
sudo service glance-api restart
sudo service nova-compute restart
sudo service cinder-volume restart
```

Once OpenStack is up and running, you should be able to create a volume with OpenStack on a Ceph block device.

BOOTING FROM A BLOCK DEVICE

If you're using OpenStack Folsom or later, you can create a volume from an image using the Cinder command line tool:

```
cinder create --image-id {id of image} --display-name {name of volume} {size of volume}
```

Note that image must be raw format. You can use qemu-img to convert from one format to another, i.e.:

```
qemu-img convert -f qcow2 -O raw precise-cloudimg.img precise-cloudimg.raw
```

Before Ceph 0.52 the image will be a full copy of the data. With Ceph 0.52 and later when Glance and Cinder are both using Ceph block devices, the image is a copy-on-write clone, so volume creation is very fast.

In the OpenStack dashboard you can then boot from that volume by launching a new instance, choosing the image that you created the volume from, and selecting ‘boot from volume’ and the volume you created.