Storage: Ceph

From Proxmox VE

Contents

Build a three node ceph storage cluster

This document explains how to setup a Ceph cluster, **storage only**, that serves Proxmox or a cluster of Proxmox. Since Proxmox 3.2, Ceph has been integrated and can be installed and configured through Proxmox specific commands, and run on Proxmox node (Proxmox nodes then act at the same time as Ceph nodes as well). For this setup, see Ceph Server

It is recommended you look through the Official installation documents for the most up to date information : http://ceph.com/docs/master/install/ For a production system you need 3 servers minimum. For testing you can get by with less, although you may be unable to properly test all the features of the cluster. Proxmox Supports CEPH >= 0.56

Prepare nodes

- Install Ubuntu.

It is recommended to use Ubuntu 12.04 LTS, this is the distribution used by Inktank for Ceph development. (you need recent filesystem version and glibc)

- Install lsb (Linux Standard Base) tool

```
apt-get install lsb-release
```

- Install XFS tools, since CEPH will use XFS

```
apt-get install xfsprogs
```

- Create SSH key on server1 and distribute it.

Generate a ssh key

```
ssh-keygen -t rsa
```

and copy it to the other servers

```
ssh-copy-id user@server2
ssh-copy-id user@server3
```

- Configure ntp on all nodes to keep time updated:

```
sudo apt-get install ntp
```

Install Ceph-Deploy

- Create entries for all other Ceph nodes in /etc/hosts
- Add Ceph repositories

```
```

```
wget -q -O- 'https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc' | sudo apt-key add -
```

- If not working try with this method

```
wget -q -O- --no-check-certificate 'https://git.ceph.com/?p=ceph.git;a=blob_plain;f=keys/release.asc' | sudo apt-key add -
```

- Add new ceph repository

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

- Install packages

```
sudo apt-get update
```

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

Create cluster using Ceph-Deploy

- Create your cluster

```
ceph-deploy new server1
```

- Install Ceph on all nodes

```
ceph-deploy install server1 server2 server3
```

You could also run:

```
ceph-deploy install server{1..3}
```

- Add a Ceph monitor.

```
ceph-deploy mon create server{1..3}
```

(You must have an odd number of monitors. If you only have one it will be a single point of failure so consider using at least 3 for high availability.)

- Gather keys

```
ceph-deploy gatherkeys server1
```

- Prepare OSDs on each server

  For each data disk, you need 1 osd daemon.

  It is assumed that these disks are empty and contain no data, zap will delete all data on disks.

  Verify the names of your data disks!

```
sudo fdisk -l
```

For servers that are not identical:

```
ceph-deploy osd --zap-disk create server1:sdb
```

```
ceph-deploy osd --zap-disk create server2:sdb
```

```
ceph-deploy osd --zap-disk create server3:sdc
```

For 3 identical servers, each with 3 data disks (sdb, sdc, sdd)

```
ceph-deploy osd --zap-disk create server{1..3}:sd{b..d}
```

By default the journal is placed on the same disk. To change this specify the path to the journal: ceph-deploy osd

prepare {node-name}:{disk}[:{path/to/journal}]

- Check the health of the cluster

```
sudo ceph -s
```

Customize Ceph

- Set your number of placement groups

```
sudo ceph osd pool set rbd pg_num 512
```

The following formula is generally used:

Total PGs = (# of OSDs * 100) / Replicas
Take this result and round up to the nearest Power of 2. For 9 OSDS you would do:
9 * 100 = 900
Default number of replicas is 2 so 900/2 = 450 rounded to the next power of 2 so 512.

- Create a new pool

```
sudo ceph osd pool create {name_of_pool} {pg_num}
```

Example:

```
sudo ceph osd pool create pve_data 512
```

- Change the number of replica groups for a pool

```
sudo ceph osd pool set {name_of_pool} size {number_of_replicas}
```

Example:

```
sudo ceph osd pool set pve_data size 3
```

Configure Proxmox to use the ceph cluster

GUI

You can use proxmox GUI to add the rbd storage

Manual configuration edit

edit your /etc/pve/storage.cfg and add the configuration

```
rbd: mycephcluster
       monhost 192.168.0.1:6789;192.168.0.2:6789;192.168.0.3:6789
       pool rbd  (optional, default =r rbd)
       username admin (optional, default = admin)
       content images
```

note: you must use ip (not dns fqdn) for monhost

Authentication

If you use cephx authentication, you need to copy the keyfile from Ceph to Proxmox VE host.

Create the /etc/pve/priv/ceph directory

```
mkdir /etc/pve/priv/ceph
```

Copy the keyring

```
scp cephserver1:/etc/ceph/ceph.client.admin.keyring /etc/pve/priv/ceph/StorageID.keyring
```

- The keyring must be named to match your Storage ID

- Copying the keyring generally requires root privileges. If you do not have the root account enabled on Ceph, you can "sudo scp" the keyring from the Ceph server to Proxmox.
- Note that for early versions of Ceph *Argonaut*, the keyring was named ceph.keyring rather than ceph.client.admin.keyring