



Ceph: Octopus + cephadm + orchestrator

Theofilos Mouratidis



Introduction

→ Ceph Octopus:

- ◆ Newest release
- ◆ Deployed with cephadm / ceph-deploy deprecated
- ◆ Goes towards a container model (docker/podman)

→ Cephadm:

- ◆ New tool to deploy clusters
- ◆ Simple to use

→ Ceph orch:

- ◆ New command to deploy a cluster

Cephadm

→ Can be either installed as a package or run as a script

```
# curl --silent --remote-name --location https://github.com/ceph/ceph/raw/octopus/src/cephadm/cephadm
# chmod +x cephadm
# dnf install -y cephadm
```

→ It bootstraps the cluster by creating a mon+mgr duo

```
# cephadm add-repo --release octopus
# cephadm install ceph-common
# mkdir -p /etc/ceph
# hostname `hostname -s`
# cephadm bootstrap --mon-ip `hostname -i`
```

→ And the result is two containers running on the main node

```
[root@octopod ~]# docker container list | grep -P "mon|mgr" --color=never
b08df12901dd      ceph/ceph:v15      "/usr/bin/ceph-mgr -..." 25 hours ago      Up 25 hours      ceph-dcd5f84a-6e87-11ea-9
a0b-fa163ec007dd-mgr.octopod.rhmnyf
3b01752818a3      ceph/ceph:v15      "/usr/bin/ceph-mon -..." 25 hours ago      Up 25 hours      ceph-dcd5f84a-6e87-11ea-9
a0b-fa163ec007dd-mon.octopod
```

Cephadm / Ceph orch: Mon

→ In case the only mon exits, it can be restarted like this:

```
# cephadm deploy -c /etc/ceph/ceph.conf --keyring /etc/ceph/ceph.client.admin.keyring --name mon.`hostname` -s` --fsid <cluster fsid>
```

→ And it creates a mon container, so the cluster can run again

→ The standard way would be to use ceph orch

```
[root@octopod ~]# ceph orch daemon restart mon.octopod
restart mon.octopod from host 'octopod'
restart mon.octopod from host 'octopod.cern.ch'
[root@octopod ~]# docker container ls
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS              NAMES
c1e9ea131ee7       ceph/ceph:v15      "/usr/bin/ceph-mon ..."  4 seconds ago      Up 4 seconds      6881/tcp           ceph-dcd5f84a-6e87-11ea-9
a0b-fa163ec007dd-mon.octopod
```

→ An entire service can also be started/stopped/restarted

◆ e.g. ceph orch restart mds

Cephadm / Ceph orch: Mon

→ Adding a new host requires two steps

- ◆ By adding the `ceph.pub` created on `/etc/ceph` to the new host
- ◆ ``ssh-copy-id -f -i ceph.pub root@<new-host>``
- ◆ ``ceph orch host add <new-host>``

→ To add a new mon:

- ◆ ``ceph orch apply mon <bootstrap-host> <host1> <host2>``
- ◆ The above command does not add a host label, it must be done explicitly
- ◆ ``ceph orch host label add <host> mon`` to filter through ``ceph orch host ls``

Ceph orchestrator: versions

- It uses cephadm by default, can also use rook
 - ◆ The tool used can be seen through `ceph orch status`
- A cluster now can be upgraded like this:
 - ◆ `ceph orch upgrade start --version 15.2.1`
 - ◆ `ceph orch upgrade status`
 - ◆ `ceph orch upgrade stop`

Ceph orchestrator: OSD

→ Once the hosts are added, we can list the available devices

```
[root@octopod ~]# ceph orch device ls --refresh
HOST      PATH      TYPE  SIZE  DEVICE      AVAIL  REJECT REASONS
octopod   /dev/vda  hdd   80.0G      False  locked
octopod   /dev/vdb  hdd   10.0G  8e62185a-9880-4f24-9  False  LVM detected, locked, Insufficient space (<5GB) on vgs
octopod   /dev/vdc  hdd   10.0G  0bc4994a-8f7a-41dd-8  False  LVM detected, locked, Insufficient space (<5GB) on vgs
octopod   /dev/vdd  hdd   10.0G  ff44ef27-92a8-4e01-8  False  LVM detected, locked, Insufficient space (<5GB) on vgs
octopod   /dev/vde  hdd   10.0G  d90affc2-7d22-478c-8  False  LVM detected, locked, Insufficient space (<5GB) on vgs
```

→ Then we can add the available devices with:

- ◆ `ceph orch apply osd --all-available-devices` or
- ◆ `ceph orch daemon add osd <host>:<device path>`

→ Or remove from the map and host with:

- ◆ `ceph orch osd rm <osd-id>` if the osd is marked as destroyed

Ceph orchestrator: OSD

→ We can also use “drivegroups.yml” to describe the devices

◆ drive_group_default:

host_pattern: '*'

data_devices:

rotational: 1

db_devices:

rotational: 0

→ `ceph orch apply osd -i /path/to/drivegroups.yml`

Ceph orchestrator: CephFS

- We can create a filesystem like this
 - ◆ ``ceph fs volume create cephfs``
 - ◆ Immediately creates an MDS container
- Now we can add more MDSes by adding more hosts
 - ◆ ``ceph orch apply mds cephfs 3 <host1> <host2> <host3>``

Conclusions

- Cephadm and ceph orchestrator simplify the creation of a CEPH cluster
- Hosts now are programmatically managed with roles
- Easy to deploy new types of daemons
- (Theoretically) version upgrade is done automatically



Thanks!

