**Ceph Install Guide – All HDD on 4.25 with comments**

To install a Ceph OSD and Monitor on 1 node. The node is called yahoo2-25. There are 12 hdd installed in the server.

Linux release: Linux yahoo2-25 3.10.0-229.4.2.el7.x86\_64 #1 SMP Wed May 13 10:06:09 UTC 2015 x86\_64 x86\_64 x86\_64 GNU/Linux

Ceph release: ceph version 0.94.2

Install steps:

1: Configure sysctl.conf: Previsioning steps

Add the following entries to the system sysctl.conf file:

net.core.somaxconn = 1024

fs.file-max = 131072

net.core.rmem\_max = 56623104

net.core.wmem\_max = 56623104

net.core.rmem\_default = 56623104

net.core.wmem\_default = 56623104

net.core.optmem\_max = 40960

net.ipv4.tcp\_rmem = 4096 87380 56623104

net.ipv4.tcp\_wmem = 4096 65536 56623104

net.core.somaxconn = 1024

net.core.netdev\_max\_backlog = 50000

net.ipv4.tcp\_max\_syn\_backlog = 30000

net.ipv4.tcp\_max\_tw\_buckets = 2000000

net.ipv4.tcp\_tw\_reuse = 1

net.ipv4.tcp\_fin\_timeout = 10

net.ipv4.tcp\_slow\_start\_after\_idle = 0

net.ipv4.udp\_rmem\_min = 8192

net.ipv4.udp\_wmem\_min = 8192

net.ipv4.conf.all.send\_redirects = 0

net.ipv4.conf.all.accept\_redirects = 0

net.ipv4.conf.all.accept\_source\_route = 0

Either reboot or issue the following statement to invoke these settings:

sysctl -p

2: Install steps for Ceph and required software:

cd /

yum install python-setuptools -y

yum install epel-release -y

yum install boost-devel\* -y

yum install easy\_install (if error, ignore)

yum install gperftools\* -y

yum install libunwind\* -y

yum install userspace\* -y

yum install lttn\* (if error, ignore)

yum install librados\* -y

yum install libceph\* -y

yum install librbd\* -y

yum install libb\* -y

easy\_install ceph-deploy

yum install yum-plugin-priorities –y

3: Modify /etc/hosts file to add hostname.

Check to see what the hostname is set to:

hostname

yahoo2-25

add to /etc/hosts: 10.241.4.25 yahoo2-25

REBOOT

4: yum install ceph\* -y All Ceph component installation

5: cd into /etc/ceph Ceph working directory

6: ceph-deploy new yahoo2-25 Ceph new monitor init

7: ceph-deploy install yahoo2-25 Install Ceph on monitor

8: ceph-deploy --overwrite-conf mon create yahoo2-25 Create Ceph monitor

9: ceph-authtool --create-keyring /tmp/ceph.mon.keyring --gen-key -n mon. --cap mon 'allow \*' Monitor keyring generate

10: ceph-authtool --create-keyring /etc/ceph/ceph.client.admin.keyring --gen-key -n client.admin --set-uid=0 --cap mon 'allow \*' --cap osd 'allow \*' --cap mds 'allow' client.admin Keyring generate

11: ceph-authtool /tmp/ceph.mon.keyring --import-keyring /etc/ceph/ceph.client.admin.keyring Combine monitor keyring with client.admin keyring

12: look into ceph.conf and grab the FSID value and change the following statement as well as the host ip address:

monmaptool --create --add yahoo2-25 10.241.4.25 --fsid a5d21036-0b15-484e-a211-d59332ffe536 /tmp/monmap --clobber Monitor map create

13: ceph-deploy --overwrite-conf admin yahoo2-25 Update monitor with new keyrings

14: service ceph restart -- should come up without OSDs Start over Ceph with all new setting

15: ceph-deploy disk list yahoo2-25 -- list of all the disks ceph finds

16: Because the disks are > 2TB, I ignored using the ceph-deploy disk zap commands.

Execute these commands if the LUNs < 2TB:

ceph-deploy disk zap yahoo2-25:sda

thru

esceph-deploy disk zap yahoo2-25:sdl

Disk zap: Init all hard disks

17: Modified ceph.conf

In /etc/ceph, a basic ceph.conf file was create during install. To add the 12 HDD to ceph, add or modify the configuration file with the following parameters: Manually add monitor and osd info in the conf file. Only 1 osd needed for experiment.

[global]

fsid = a5d21036-0b15-484e-a211-d59332ffe536

mon\_initial\_members = yahoo2-25

mon\_host = 10.241.4.25

auth\_cluster\_required = cephx

auth\_service\_required = cephx

auth\_client\_required = cephx

auth supported = none

filestore\_xattr\_use\_omap = true

[osd.0]

host = yahoo2-25

[osd.1]

host = yahoo2-25

[osd.2]

host = yahoo2-25

[osd.3]

host = yahoo2-25

[osd.4]

host = yahoo2-25

[osd.5]

host = yahoo2-25

[osd.6]

host = yahoo2-25

[osd.7]

host = yahoo2-25

[osd.8]

host = yahoo2-25

[osd.9]

host = yahoo2-25

[osd.10]

host = yahoo2-25

[osd.11]

host = yahoo2-25

[mon.yahoo2-25]

host = yahoo2-25

mon\_addr = 10.241.4.25:6789

18: create 12 OSDs from 0 to 11: Create osds. Note: only 1 osd needed for experiment.

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

ceph osd create

19: Make filesystems for each lun: OSD’s hard disk file system init

mkfs.xfs /dev/sda -f

mkfs.xfs /dev/sdb -f

mkfs.xfs /dev/sdc -f

mkfs.xfs /dev/sdd -f

mkfs.xfs /dev/sde -f

mkfs.xfs /dev/sdf -f

mkfs.xfs /dev/sdg -f

mkfs.xfs /dev/sdh -f

mkfs.xfs /dev/sdi -f

mkfs.xfs /dev/sdj -f

mkfs.xfs /dev/sdk -f

mkfs.xfs /dev/sdl -f

20: Make mountpoint dirs. for each HDD: Make mount point on EXTRA DISK (Not on the OS disk)

mkdir /var/lib/ceph/osd/ceph-0

mkdir /var/lib/ceph/osd/ceph-1

mkdir /var/lib/ceph/osd/ceph-2

mkdir /var/lib/ceph/osd/ceph-3

mkdir /var/lib/ceph/osd/ceph-4

mkdir /var/lib/ceph/osd/ceph-5

mkdir /var/lib/ceph/osd/ceph-6

mkdir /var/lib/ceph/osd/ceph-7

mkdir /var/lib/ceph/osd/ceph-8

mkdir /var/lib/ceph/osd/ceph-9

mkdir /var/lib/ceph/osd/ceph-10

mkdir /var/lib/ceph/osd/ceph-11

21: Mount LUNs: Mount the EXTRA hard disks

mount /dev/sda /var/lib/ceph/osd/ceph-0

mount /dev/sdb /var/lib/ceph/osd/ceph-1

mount /dev/sdc /var/lib/ceph/osd/ceph-2

mount /dev/sdd /var/lib/ceph/osd/ceph-3

mount /dev/sde /var/lib/ceph/osd/ceph-4

mount /dev/sdf /var/lib/ceph/osd/ceph-5

mount /dev/sdg /var/lib/ceph/osd/ceph-6

mount /dev/sdh /var/lib/ceph/osd/ceph-7

mount /dev/sdi /var/lib/ceph/osd/ceph-8

mount /dev/sdj /var/lib/ceph/osd/ceph-9

mount /dev/sdk /var/lib/ceph/osd/ceph-10

mount /dev/sdl /var/lib/ceph/osd/ceph-11

\*\*\*\*\* modify /etc/fstab with the proceeding mount info

22: Create keyring for each OSD using ceph.conf fsid: Generate osd’s keyring with this Ceph’s fsid(uuid)

ceph-osd -i 0 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 1 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 2 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 3 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 4 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 5 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 6 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 7 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 8 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 9 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 10 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

ceph-osd -i 11 --mkfs --mkkey --osd-uuid 34a04b03-8925-4cf4-915d-aafab58d7d7d

23: Delete Ceph authourization for each OSD: Remove default osd’s authorization

ceph auth del osd.0

ceph auth del osd.1

ceph auth del osd.2

ceph auth del osd.3

ceph auth del osd.4

ceph auth del osd.5

ceph auth del osd.6

ceph auth del osd.7

ceph auth del osd.8

ceph auth del osd.9

ceph auth del osd.10

ceph auth del osd.11

24: Add authourization to each OSD: Give all-pass authorization on the osd

ceph auth add osd.0 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-0/keyring

ceph auth add osd.1 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-1/keyring

ceph auth add osd.2 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-2/keyring

ceph auth add osd.3 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-3/keyring

ceph auth add osd.4 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-4/keyring

ceph auth add osd.5 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-5/keyring

ceph auth add osd.6 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-6/keyring

ceph auth add osd.7 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-7/keyring

ceph auth add osd.8 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-8/keyring

ceph auth add osd.9 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-9/keyring

ceph auth add osd.10 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-10/keyring

ceph auth add osd.11 osd 'allow \*' mon 'allow profile osd' -i /var/lib/ceph/osd/ceph-11/keyring

25: ceph osd crush add-bucket yahoo2-25 host Add Ceph node to CRUSH map

26: ceph osd crush move yahoo2-25 root=default Place Ceph node under root default

27: ceph osd crush add osd.0 1.0 host=yahoo2-25 Add osd node to CRUSH map, so it can begin receiving data.

ceph osd crush add osd.1 1.0 host=yahoo2-25

ceph osd crush add osd.2 1.0 host=yahoo2-25

ceph osd crush add osd.3 1.0 host=yahoo2-25

ceph osd crush add osd.4 1.0 host=yahoo2-25

ceph osd crush add osd.5 1.0 host=yahoo2-25

ceph osd crush add osd.6 1.0 host=yahoo2-25

ceph osd crush add osd.7 1.0 host=yahoo2-25

ceph osd crush add osd.8 1.0 host=yahoo2-25

ceph osd crush add osd.9 1.0 host=yahoo2-25

ceph osd crush add osd.10 1.0 host=yahoo2-25

ceph osd crush add osd.11 1.0 host=yahoo2-25

28: Start each OSD: Start osd service

/etc/init.d/ceph start osd.0

/etc/init.d/ceph start osd.1

/etc/init.d/ceph start osd.2

/etc/init.d/ceph start osd.3

/etc/init.d/ceph start osd.4

/etc/init.d/ceph start osd.5

/etc/init.d/ceph start osd.6

/etc/init.d/ceph start osd.7

/etc/init.d/ceph start osd.8

/etc/init.d/ceph start osd.9

/etc/init.d/ceph start osd.10

/etc/init.d/ceph start osd.11

29: Indicate that all OSD task are performed: Necessary file create needed

touch /var/lib/ceph/osd/ceph-0/sysvinit

touch /var/lib/ceph/osd/ceph-1/sysvinit

touch /var/lib/ceph/osd/ceph-2/sysvinit

touch /var/lib/ceph/osd/ceph-3/sysvinit

touch /var/lib/ceph/osd/ceph-4/sysvinit

touch /var/lib/ceph/osd/ceph-5/sysvinit

touch /var/lib/ceph/osd/ceph-6/sysvinit

touch /var/lib/ceph/osd/ceph-7/sysvinit

touch /var/lib/ceph/osd/ceph-8/sysvinit

touch /var/lib/ceph/osd/ceph-9/sysvinit

touch /var/lib/ceph/osd/ceph-10/sysvinit

touch /var/lib/ceph/osd/ceph-11/sysvinit

30: ceph -s Ceph status check

cluster a5d21036-0b15-484e-a211-d59332ffe536

health HEALTH\_WARN

64 pgs degraded

64 pgs stuck degraded

64 pgs stuck inactive

64 pgs stuck unclean

64 pgs stuck undersized

64 pgs undersized

too few PGs per OSD (5 < min 30)

monmap e1: 1 mons at {yahoo2-25=10.241.4.25:6789/0}

election epoch 2, quorum 0 yahoo2-25

osdmap e37: 12 osds: 12 up, 12 in

pgmap v61: 64 pgs, 1 pools, 0 bytes data, 0 objects

61834 MB used, 44629 GB / 44690 GB avail

64 undersized+degraded+peered

(could display warning messages at this point)

31: service ceph restart

35: ceph health

HEALTH\_WARN 64 pgs degraded; 64 pgs stuck degraded; 64 pgs stuck inactive; 64 pgs stuck unclean; 64 pgs stuck undersized; 64 pgs undersized; too few PGs per OSD (5 < min 30)

**To benchmark the Ceph Storage Cluster using RADOS**

Each of the 3 tests below will run for 500 seconds with a default thread count of 16.

1: ceph osd pool create pool1 256 256

2: ceph osd pool set pool1 size 1

3: rados bench -p pool1 500 write --no-cleanup RADOS write bench

(no cleanup allows to use the data for reads)

3: rados bench -p pool1 500 seq (seq reads) RADOS sequential read bench

4: rados bench -p pool1 500 rand (random reads) RADOS random read bench

5: rados -p pool1 cleanup (cleanup the data) Clear cache