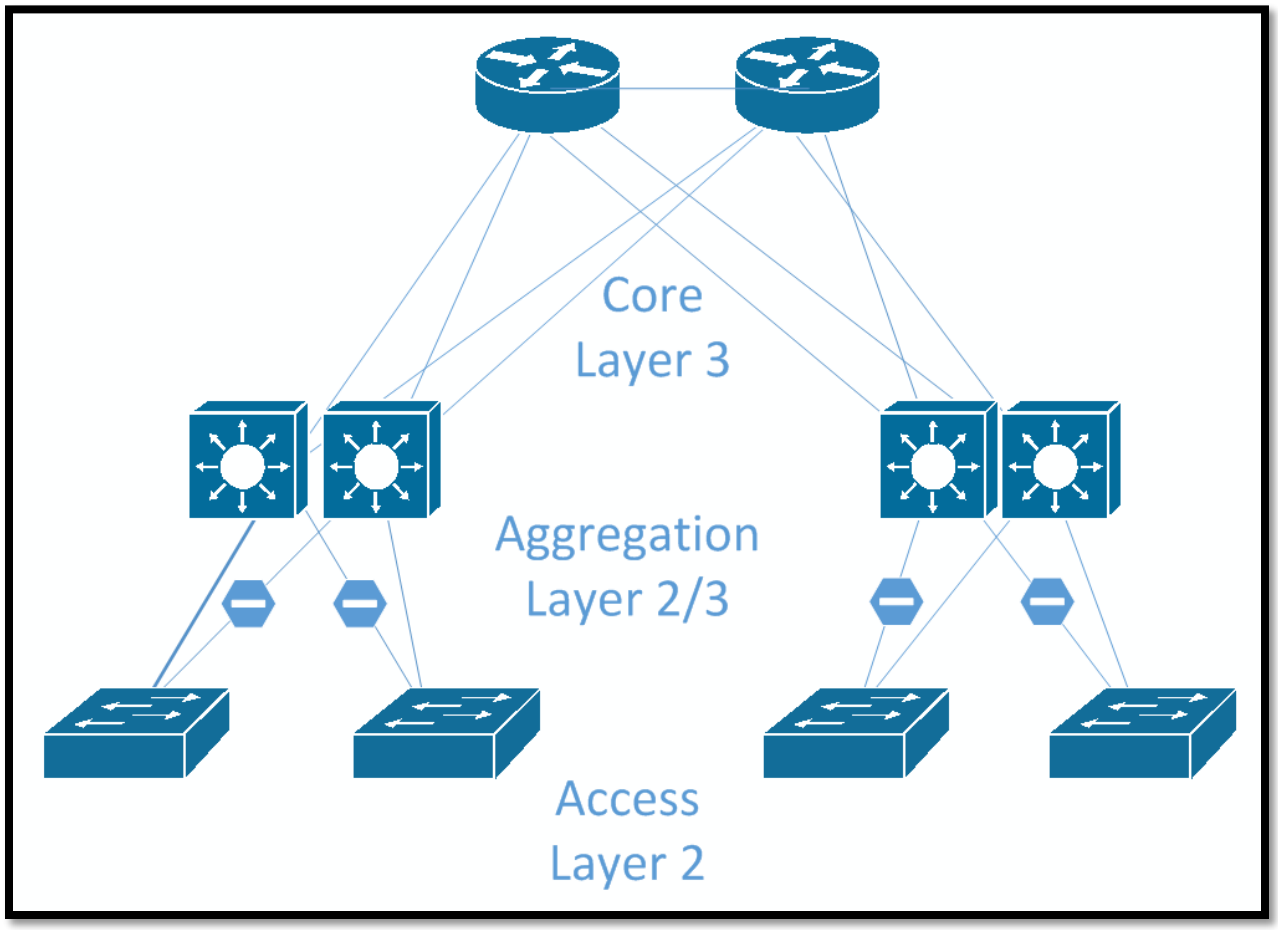
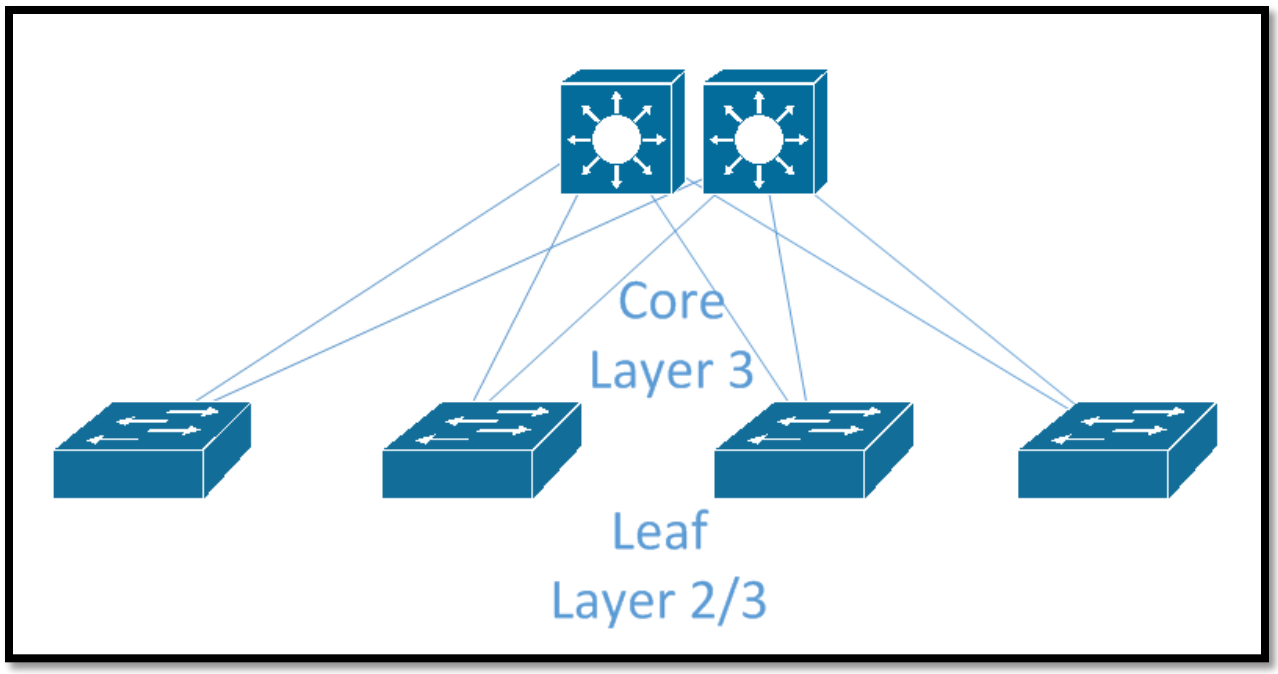
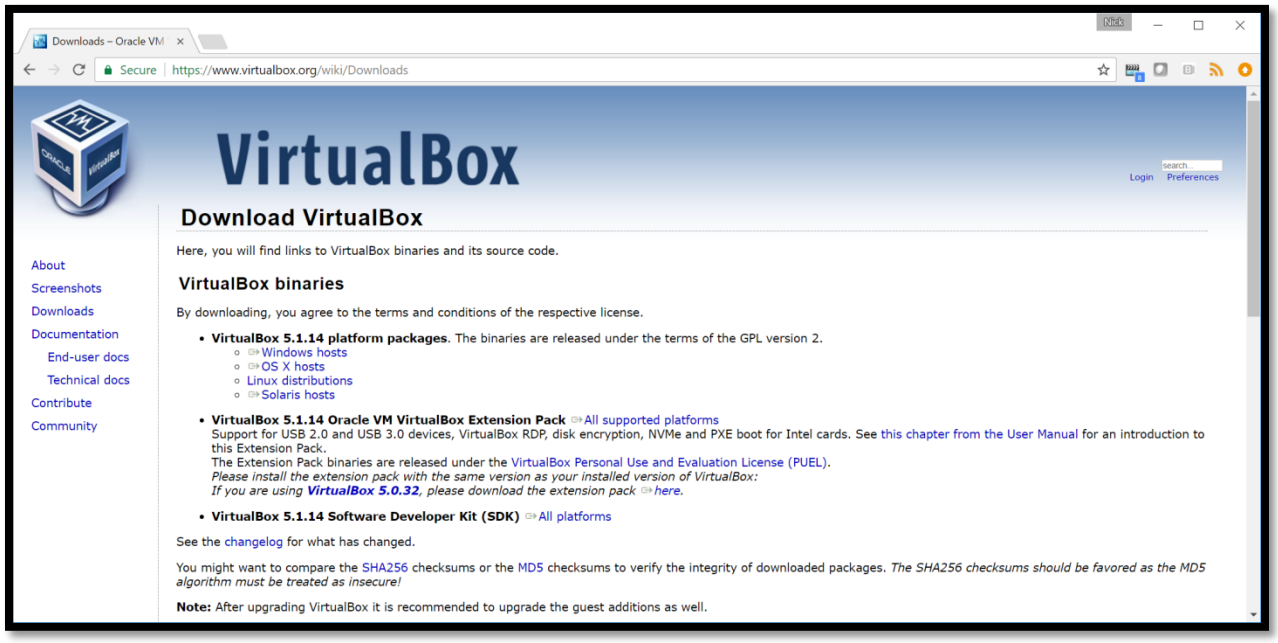
Chapter 1: Planning for Ceph




Core
Layer 3
Leaf
Layer 2/3

Chapter 2: Deploying Ceph

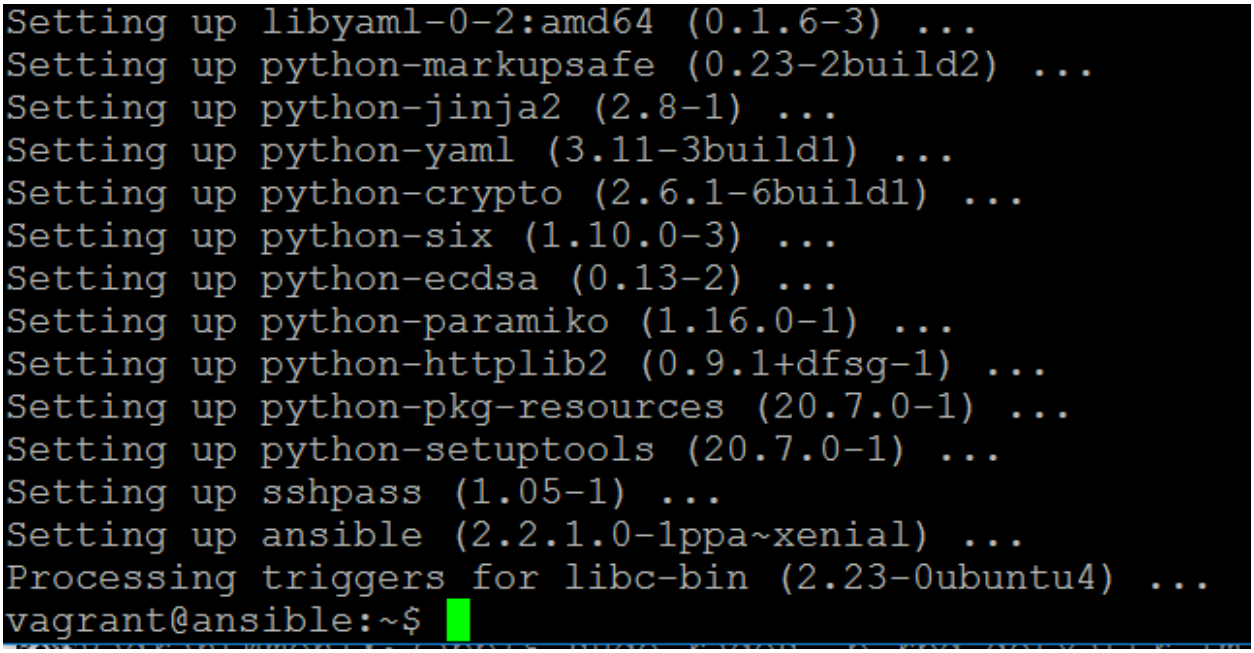



```
C:\Users\nfisk\vagrant>cd ceph

C:\Users\nfisk\vagrant\ceph>
```

```
Installing the 'vagrant-hostmanager' plugin. This can take a few minutes...
Fetching: vagrant-hostmanager-1.8.5.gem (100%)
Installed the plugin 'vagrant-hostmanager (1.8.5)'!

==> box: Loading metadata for box 'bento/ubuntu-16.04'
    box: URL: https://atlas.hashicorp.com/bento/ubuntu-16.04
This box can work with multiple providers! The providers that it
can work with are listed below. Please review the list and choose
the provider you will be working with.

1) parallels
2) virtualbox
3) vmware_desktop

Enter your choice: 2
==> box: Adding box 'bento/ubuntu-16.04' (v2.3.1) for provider: virtualbox
    box: Downloading: https://atlas.hashicorp.com/bento/boxes/ubuntu-16.04/versions/2.3.1/providers/virtualbox.box
    box: Progress: 100% (Rate: 5257k/s, Estimated time remaining: --:--:--)
==> box: Successfully added box 'bento/ubuntu-16.04' (v2.3.1) for 'virtualbox'!
Bringing machine 'ansible' up with 'virtualbox' provider...
Bringing machine 'mon1' up with 'virtualbox' provider...
Bringing machine 'mon2' up with 'virtualbox' provider...
Bringing machine 'mon3' up with 'virtualbox' provider...
Bringing machine 'osd1' up with 'virtualbox' provider...
Bringing machine 'osd2' up with 'virtualbox' provider...
Bringing machine 'osd3' up with 'virtualbox' provider...
==> ansible: Importing base box 'bento/ubuntu-16.04'...
==> ansible: Matching MAC address for NAT networking...
==> ansible: Checking if box 'bento/ubuntu-16.04' is up to date...
==> ansible: Setting the name of the VM: ceph_ansible_1486503043550_56998
==> ansible: Clearing any previously set network interfaces...
==> ansible: Preparing network interfaces based on configuration...
    ansible: Adapter 1: nat
    ansible: Adapter 2: hostonly
==> ansible: Forwarding ports...
    ansible: 22 (guest) => 2222 (host) (adapter 1)
==> ansible: Running 'pre-boot' VM customizations...
==> ansible: Booting VM...
==> ansible: Waiting for machine to boot. This may take a few minutes...
```

```
`ssh` executable not found in any directories in the %PATH% variable. Is an
SSH client installed? Try installing Cygwin, MinGW or Git, all of which
contain an SSH client. Or use your favorite SSH client with the following
authentication information shown below:

Host: 127.0.0.1
Port: 2200
Username: vagrant
login as: vagrant
vagrant@127.0.0.1's password:
Welcome to Ubuntu 16.04.1 LTS (GNU/Linux 4.4.0-51-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

0 packages can be updated.
0 updates are security updates.


vagrant@ansible:~$
```

```
 Ansible is a radically simple IT automation platform that makes your applications and systems easie
r to deploy. Avoid writing scripts or custom code to deploy and update your applications— automate i
n a language that approaches plain English, using SSH, with no agents to install on remote systems.

http://ansible.com/
 More info: https://launchpad.net/~ansible/+archive/ubuntu/ansible
Press [ENTER] to continue or ctrl-c to cancel adding it

gpg: keyring `/tmp/tmpt5a6qdao/secring.gpg' created
gpg: keyring `/tmp/tmpt5a6qdao/pubring.gpg' created
gpg: requesting key 7BB9C367 from hkp server keyserver.ubuntu.com
gpg: /tmp/tmpt5a6qdao/trustdb.gpg: trustdb created
gpg: key 7BB9C367: public key "Launchpad PPA for Ansible, Inc." imported
gpg: Total number processed: 1
gpg:               imported: 1  (RSA: 1)
OK
```


Setting up libyaml-0-2:amd64 (0.1.6-3) ...
Setting up python-markupsafe (0.23-2build2) ...
Setting up python-jinja2 (2.8-1) ...
Setting up python-yaml (3.11-3build1) ...
Setting up python-crypto (2.6.1-6build1) ...
Setting up python-six (1.10.0-3) ...
Setting up python-ecdsa (0.13-2) ...
Setting up python-paramiko (1.16.0-1) ...
Setting up python-httplib2 (0.9.1+dfsg-1) ...
Setting up python-pkg-resources (20.7.0-1) ...
Setting up python-setuptools (20.7.0-1) ...
Setting up sshpass (1.05-1) ...
Setting up ansible (2.2.1.0-1ppa~xenial) ...
Processing triggers for libc-bin (2.23-0ubuntu4) ...
vagrant@ansible:~$

```
vagrant@ansible:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/vagrant/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/vagrant/.ssh/id_rsa.
Your public key has been saved in /home/vagrant/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:mdvKrx6ZG88AKQsPnaFpjKlPb8pmmnfqDiQPv4OQnpw vagrant@ansible
The key's randomart image is:
+---[RSA 2048]----+
|                 |
|                 |
|    .            |
| + + o . o       |
|=oB + o S        |
|*= + o . =       |
|*.* o   B .      |
|.E=oo  . O       |
|oBO*.  .*o+      |
+----[SHA256]-----+
vagrant@ansible:~$ ssh-copy-id mon1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/vagrant/.ssh/id_rsa.pub"
The authenticity of host 'mon1 (192.168.0.41)' can't be established.
ECDSA key fingerprint is SHA256:RI5/3ep65qXeDkZSACi/rN0hBxiLrBxMvcyk9CfLkyg.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are alr
eady installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to inst
all the new keys
vagrant@mon1's password:

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'mon1'"
and check to make sure that only the key(s) you wanted were added.
```

```
vagrant@ansible:~$ ssh mon1
Welcome to Ubuntu 16.04.1 LTS (GNU/Linux 4.4.0-51-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

0 packages can be updated.
0 updates are security updates.


vagrant@mon1:~$
```

```
vagrant@ansible:~$ ansible mon1 -m ping
mon1 | SUCCESS => {
    "changed": false,
    "ping": "pong"
}
vagrant@ansible:~$
```

```
vagrant@ansible:~$ ansible mon1 -a 'uname -r'
mon1 | SUCCESS | rc=0 >>
4.4.0-51-generic

vagrant@ansible:~$
```

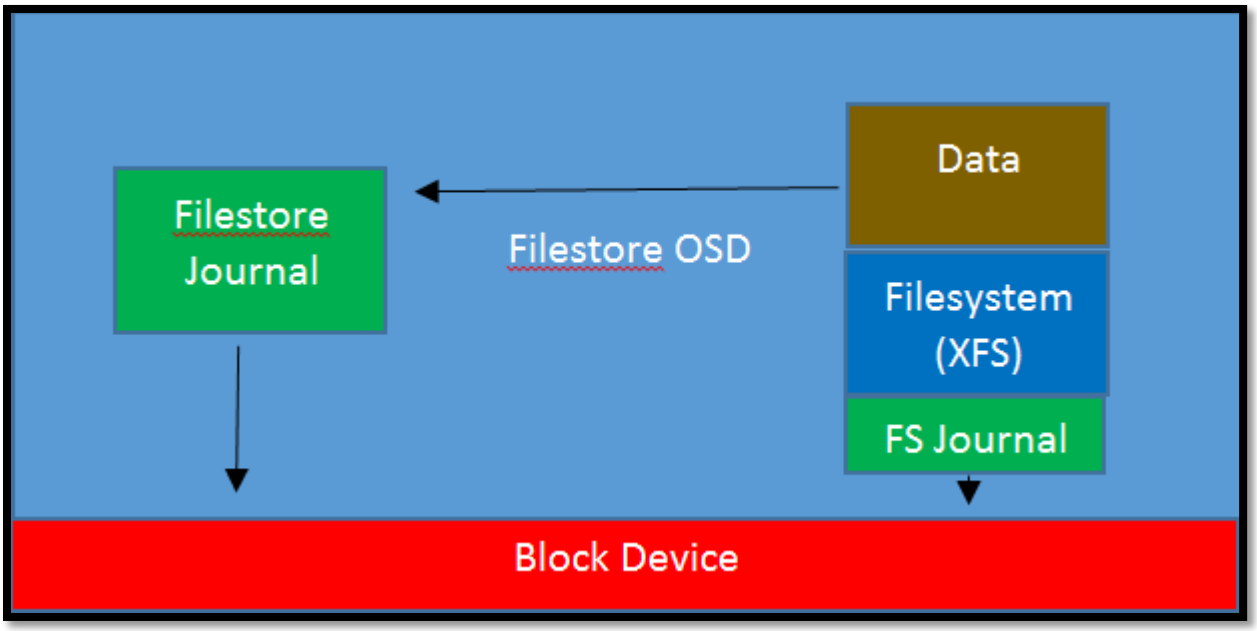
```
vagrant@ansible:~$ ansible-playbook /etc/ansible/playbook.yml

PLAY [mon1 osd1] ***************************************************************

TASK [setup] *******************************************************************
ok: [mon1]
ok: [osd1]

TASK [Echo Variables] **********************************************************
ok: [mon1] => {
    "msg": "I am a foo"
}
ok: [osd1] => {
    "msg": "I am a bar"
}

PLAY RECAP *********************************************************************
mon1                       : ok=2    changed=0    unreachable=0    failed=0
osd1                       : ok=2    changed=0    unreachable=0    failed=0

vagrant@ansible:~$
```

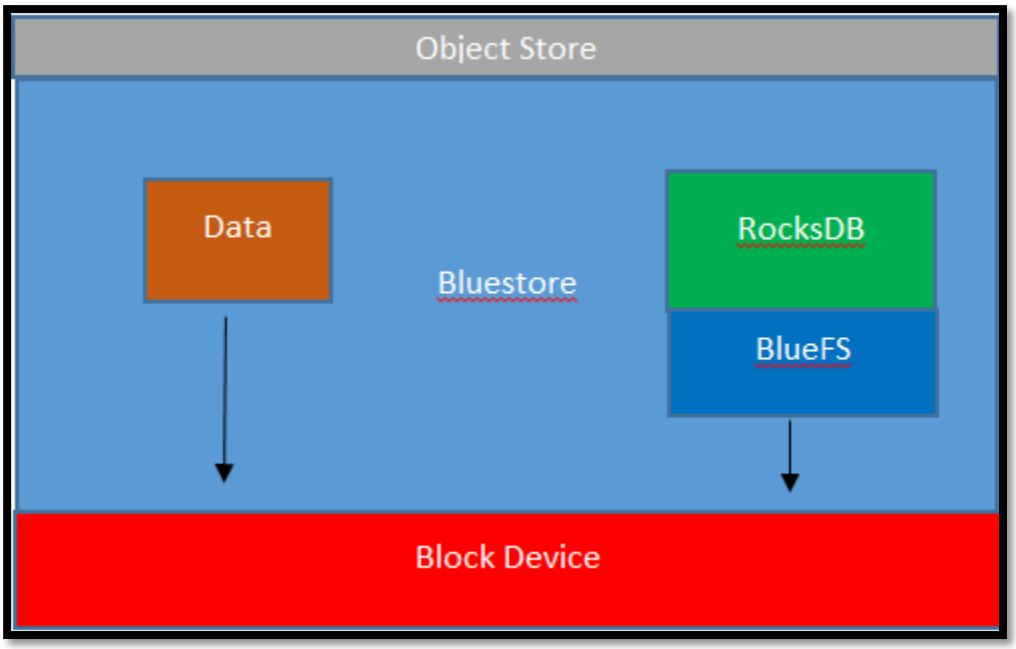
```
vagrant@ansible:~$ git clone https://github.com/ceph/ceph-ansible.git
Cloning into 'ceph-ansible'...
remote: Counting objects: 13875, done.
remote: Compressing objects: 100% (69/69), done.
remote: Total 13875 (delta 32), reused 0 (delta 0), pack-reused 13802
Receiving objects: 100% (13875/13875), 2.29 MiB | 1.94 MiB/s, done.
Resolving deltas: 100% (9234/9234), done.
Checking connectivity... done.
vagrant@ansible:~$ sudo cp -a ceph-ansible/* /etc/ansible/
vagrant@ansible:~$
```

```
PLAY RECAP *********************************************************************
mon1                       : ok=57   changed=15   unreachable=0    failed=0
mon2                       : ok=51   changed=12   unreachable=0    failed=0
mon3                       : ok=51   changed=12   unreachable=0    failed=0
osd1                       : ok=59   changed=11   unreachable=0    failed=0
osd2                       : ok=57   changed=11   unreachable=0    failed=0
osd3                       : ok=57   changed=11   unreachable=0    failed=0
```

```
vagrant@ansible:/etc/ansible$ ssh mon1
Welcome to Ubuntu 16.04.1 LTS (GNU/Linux 4.4.0-51-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

93 packages can be updated.
28 updates are security updates.


Last login: Tue Feb  7 22:08:42 2017 from 192.168.0.40
vagrant@mon1:~$ sudo ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_OK
     monmap e1: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 6, quorum 0,1,2 mon1,mon2,mon3
     osdmap e8: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v15: 64 pgs, 1 pools, 0 bytes data, 0 objects
            100 MB used, 26794 MB / 26894 MB avail
                  64 active+clean
```

Chapter 3:Bluestore






```
vagrant@mon1:~$ sudo ceph osd out 2
marked out osd.2.
```

```
vagrant@mon1:~$ sudo ceph osd crush remove osd.2
removed item id 2 name 'osd.2' from crush map
```

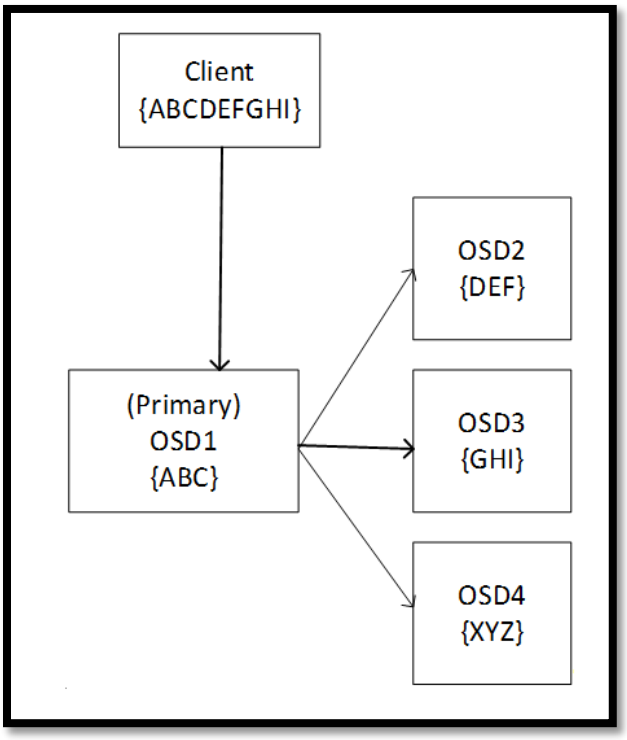
```
vagrant@mon1:~$ sudo ceph auth del osd.2
updated
```

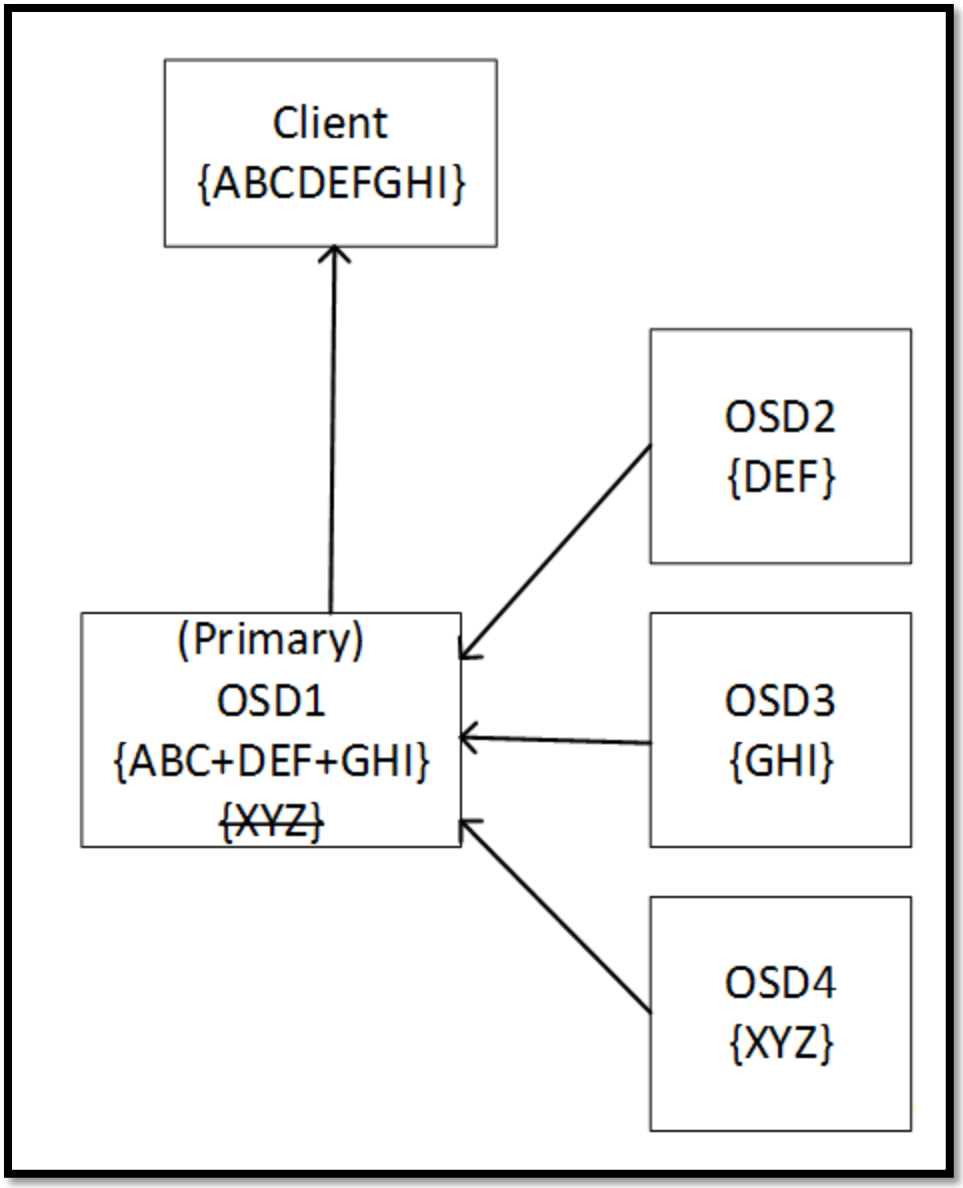
```
vagrant@mon1:~$ sudo ceph osd rm osd.2
removed osd.2
```

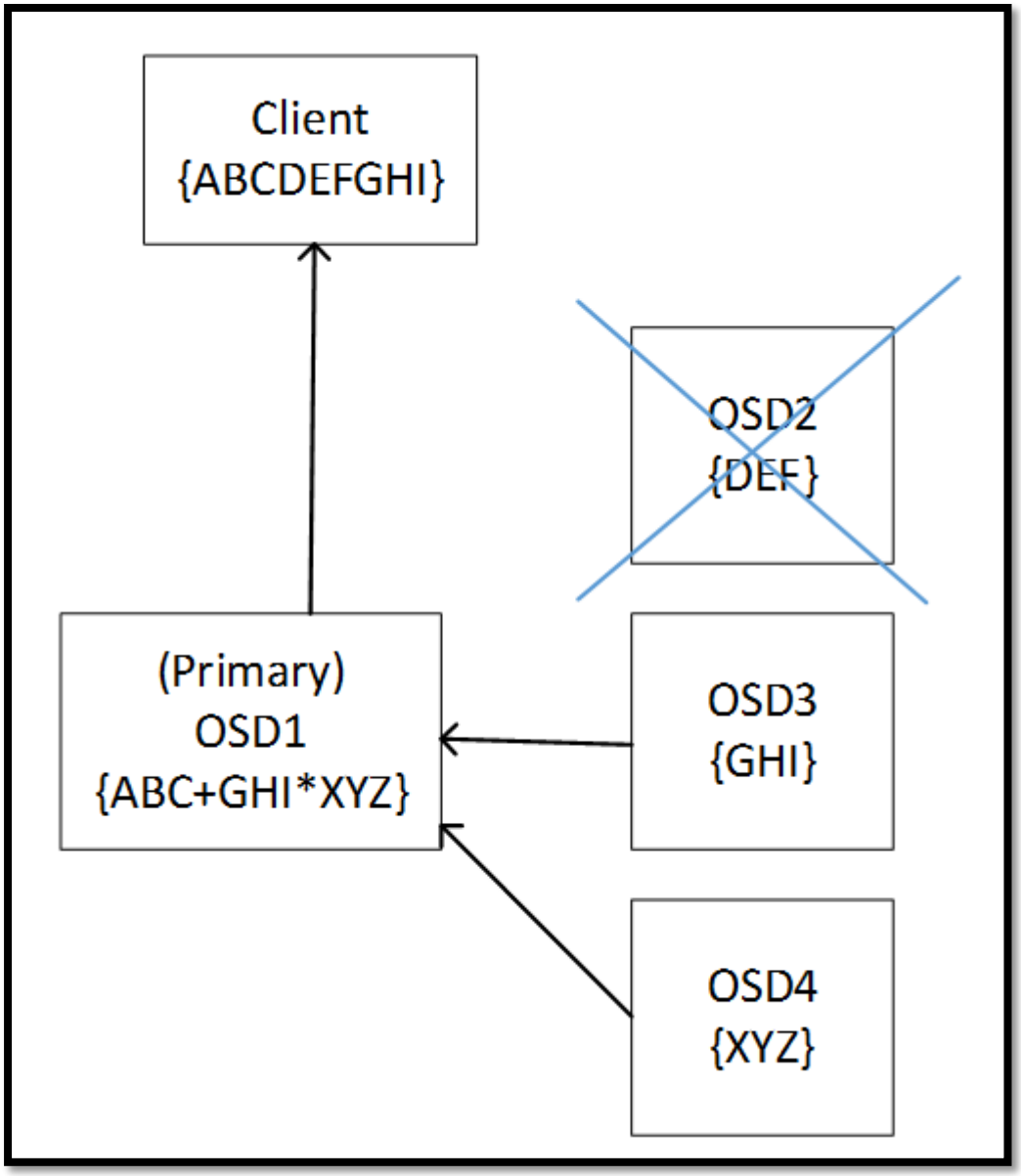
```
vagrant@osd3:~$ sudo ceph-disk zap /dev/sdb
Caution: invalid backup GPT header, but valid main header; regenerating
backup header from main header.

Warning! Main and backup partition tables differ! Use the 'c' and 'e' options
on the recovery & transformation menu to examine the two tables.

Warning! One or more CRCs don't match. You should repair the disk!

********************************************************************************
Caution: Found protective or hybrid MBR and corrupt GPT. Using GPT, but disk
verification and recovery are STRONGLY recommended.
********************************************************************************
GPT data structures destroyed! You may now partition the disk using fdisk or
other utilities.
Creating new GPT entries.
The operation has completed successfully.
```

```
vagrant@osd3:~$ sudo ceph-disk prepare --bluestore /dev/sdb
Setting name!
partNum is 0
REALLY setting name!
The operation has completed successfully.
Setting name!
partNum is 1
REALLY setting name!
The operation has completed successfully.
The operation has completed successfully.
meta-data=/dev/sdb1              isize=2048   agcount=4, agsize=6400 blks
         =                       sectsz=512   attr=2, projid32bit=1
         =                       crc=1        finobt=1, sparse=0
data     =                       bsize=4096   blocks=25600, imaxpct=25
         =                       sunit=0      swidth=0 blks
naming   =version 2              bsize=4096   ascii-ci=0 ftype=1
log      =internal log           bsize=4096   blocks=864, version=2
         =                       sectsz=512   sunit=0 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
The operation has completed successfully.
```

Chapter 4: Erasure coding for better storage efficiency




Client
{ABCDEFGHI}
OSD2
{DEF}
(Primary)
OSD1
{ABC+DEF+GHI}
~~{XYZ}~~
OSD3
{GHI}
OSD4
{XYZ}




```
vagrant@mon1:~$ sudo ceph osd erasure-code-profile ls
default
```

```
vagrant@mon1:~$ sudo ceph osd erasure-code-profile get default
k=2
m=1
plugin=jerasure
technique=reed_sol_van
```

```
vagrant@mon1:~$ sudo ceph osd erasure-code-profile ls
default
example_profile
```

```
vagrant@mon1:~$ sudo ceph osd pool create ecpool 128 128 erasure example_profile
pool 'ecpool' created
```

```
root@mon1:/home/vagrant# echo "I am test data for a test object" | rados --pool ecpool put Test1 -
root@mon1:/home/vagrant# rados --pool ecpool get Test1 -
I am test data for a test object
```

```
root@mon1:/home/vagrant# ceph osd map ecpool Test1
osdmap e114 pool 'ecpool' (3) object 'Test1' -> pg 3.ae48bdc0 (3.40) -> up ([1,2,0], p1) acting ([1,2,0
], p1)
```

```
root@osd1:/home/vagrant# ls -l /var/lib/ceph/osd/ceph-0/current/3.40s2_head/
total 4
-rw-r--r-- 1 ceph ceph    0 Feb 12 19:53 __head_00000040__3_ffffffffffffffff_2
-rw-r--r-- 1 ceph ceph 2048 Feb 12 19:56 Test1__head_AE48BDC0__3_ffffffffffffffff_2
```

```
root@osd2:/home/vagrant# ls -l /var/lib/ceph/osd/ceph-2/current/3.40s1_head/
total 4
-rw-r--r-- 1 ceph ceph    0 Feb 12 19:53 __head_00000040__3_ffffffffffffffff_1
-rw-r--r-- 1 ceph ceph 2048 Feb 12 19:56 Test1__head_AE48BDC0__3_ffffffffffffffff_1
```

```
root@osd3:/home/vagrant# ls -l /var/lib/ceph/osd/ceph-1/current/3.40s0_head/
total 4
-rw-r--r-- 1 ceph ceph    0 Feb 12 19:53 __head_00000040__3_ffffffffffffffff_0
-rw-r--r-- 1 ceph ceph 2048 Feb 12 19:56 Test1__head_AE48BDC0__3_ffffffffffffffff_0
```

```
vagrant@ansible:/etc/ansible$ ansible-playbook -K infrastructure-playbooks/rolling_update.yml
SUDO password:
[DEPRECATION WARNING]: docker is kept for backwards compatibility but usage is discouraged. The module
documentation details page may explain more about this rationale..
This feature will be removed in a
future release. Deprecation warnings can be disabled by setting deprecation_warnings=False in
ansible.cfg.
Are you sure you want to upgrade the cluster? [no]: yes
```

```
PLAY RECAP *********************************************************************
localhost                  : ok=1    changed=0    unreachable=0    failed=0
mon1                       : ok=73   changed=13   unreachable=0    failed=0
mon2                       : ok=68   changed=7    unreachable=0    failed=0
mon3                       : ok=68   changed=7    unreachable=0    failed=0
osd1                       : ok=69   changed=9    unreachable=0    failed=0
osd2                       : ok=69   changed=9    unreachable=0    failed=0
osd3                       : ok=69   changed=9    unreachable=0    failed=0
```

```
vagrant@mon1:~$ ceph -v
ceph version 11.2.0 (f223e27eeb35991352ebc1f67423d4ebc252adb7)
```

```
vagrant@mon1:~$ sudo ceph -s
2017-02-10 20:56:29.825996 7f6f18fc9700 -1 WARNING: the following dangerous and experimental features are enabled: bluestor
e,debug_white_box_testing_ec_overwrites
2017-02-10 20:56:29.831159 7f6f18fc9700 -1 WARNING: the following dangerous and experimental features are enabled: bluestor
e,debug_white_box_testing_ec_overwrites
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_WARN
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 46, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon1 standbys: mon2, mon3
     osdmap e74: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v600: 64 pgs, 1 pools, 3920 bytes data, 2 objects
            107 MB used, 26787 MB / 26894 MB avail
                  64 active+clean
```

```
pg 2.7a is creating+incomplete, acting [0,2,1,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.79 is creating+incomplete, acting [1,0,2,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.78 is creating+incomplete, acting [1,0,2147483647,2] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.7f is creating+incomplete, acting [0,2,1,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
```

```
vagrant@mon1:~$ sudo ceph osd pool create broken_ecpool 128 128 erasure broken_profile
2017-02-12 19:25:55.660243 7f3c6b74e700 -1 WARNING: the following dangerous and experimental features a
re enabled: bluestore,debug_white_box_testing_ec_overwrites
2017-02-12 19:25:55.671201 7f3c6b74e700 -1 WARNING: the following dangerous and experimental features a
re enabled: bluestore,debug_white_box_testing_ec_overwrites
pool 'broken_ecpool' created
```

```
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_ERR
            128 pgs are stuck inactive for more than 300 seconds
            128 pgs incomplete
            128 pgs stuck inactive
            128 pgs stuck unclean
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 64, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon1 standbys: mon2, mon3
     osdmap e98: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v695: 192 pgs, 2 pools, 3920 bytes data, 2 objects
            112 MB used, 26782 MB / 26894 MB avail
                 128 creating+incomplete
                  64 active+clean
```


```
pg 2.7a is creating+incomplete, acting [0,2,1,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.79 is creating+incomplete, acting [1,0,2,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.78 is creating+incomplete, acting [1,0,2147483647,2] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
pg 2.7f is creating+incomplete, acting [0,2,1,2147483647] (reducing pool broken_ecpool min_size from 4
may help; search ceph.com/docs for 'incomplete')
```

Chapter 5: Deployiong with Librados

```
vagrant@mon1:~$ sudo apt-get install build-essential
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  dpkg-dev g++ g++-5 libalgorithm-diff-perl libalgorithm-diff-xs-perl libalgorithm-merge-perl
  libstdc++-5-dev
Suggested packages:
  debian-keyring g++-multilib g++-5-multilib gcc-5-doc libstdc++6-5-dbg libstdc++-5-doc
The following NEW packages will be installed:
  build-essential dpkg-dev g++ g++-5 libalgorithm-diff-perl libalgorithm-diff-xs-perl
  libalgorithm-merge-perl libstdc++-5-dev
0 upgraded, 8 newly installed, 0 to remove and 93 not upgraded.
Need to get 10.4 MB of archives.
After this operation, 41.0 MB of additional disk space will be used.
Do you want to continue? [Y/n]
```

```
vagrant@mon1:~$ sudo apt-get install librados-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  librados-dev
0 upgraded, 1 newly installed, 0 to remove and 93 not upgraded.
Need to get 42.0 MB of archives.
After this operation, 358 MB of additional disk space will be used.
Get:1 http://download.ceph.com/debian-jewel xenial/main amd64 librados-dev amd64 10.2.5-1xenial [42.0 MB]
Fetched 42.0 MB in 30s (1,359 kB/s)
Selecting previously unselected package librados-dev.
(Reading database ... 40080 files and directories currently installed.)
Preparing to unpack .../librados-dev_10.2.5-1xenial_amd64.deb ...
Unpacking librados-dev (10.2.5-1xenial) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up librados-dev (10.2.5-1xenial) ...
```

```
vagrant@mon1:~/test_app$ sudo ./test_app
RADOS initialised
Ceph config parsed
Connected to the rados cluster
RADOS connection destroyed
The END
```

```
Reading package lists... Done
Building dependency tree
Reading state information... Done
python-rados is already the newest version (10.2.5-1xenial).
python-rados set to manually installed.
The following additional packages will be installed:
  libjbig0 libjpeg-turbo8 libjpeg8 liblcms2-2 libtiff5 libwebp5 libwebpmux1 python-pil
Suggested packages:
  liblcms2-utils python-pil-doc python-pil-dbg
The following NEW packages will be installed:
  libjbig0 libjpeg-turbo8 libjpeg8 liblcms2-2 libtiff5 libwebp5 libwebpmux1 python-imaging python-pil
0 upgraded, 9 newly installed, 0 to remove and 93 not upgraded.
Need to get 916 kB of archives.
After this operation, 3,303 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

```
vagrant@mon1:~$ sudo python app1.py --help
usage: app1.py [-h] --action ACTION --image-file IMAGEFILE --object-name
               OBJECTNAME --pool POOL [--comment COMMENT]

Image to RADOS Object Utility

optional arguments:
  -h, --help            show this help message and exit
  --action ACTION       Either upload or download image to/from Ceph
  --image-file IMAGEFILE
                        The image file to upload to RADOS
  --object-name OBJECTNAME
                        The name of the RADOS object
  --pool POOL           The name of the RADOS pool to store the object
  --comment COMMENT     A comment to store with the object
```

```
vagrant@mon1:~$ wget http://docs.ceph.com/docs/master/_static/logo.png
--2017-02-08 20:37:01--  http://docs.ceph.com/docs/master/_static/logo.png
Resolving docs.ceph.com (docs.ceph.com)... 158.69.67.53
Connecting to docs.ceph.com (docs.ceph.com)|158.69.67.53|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3898 (3.8K) [image/png]
Saving to: 'logo.png'

logo.png                 100%[=====================================>]   3.81K  --.-KB/s    in 0s

2017-02-08 20:37:01 (106 MB/s) - 'logo.png' saved [3898/3898]
```

```
vagrant@mon1:~$ sudo python app1.py --action=upload --image-file=logo.png --object-name=image_test --pool
=rbd --comment="Ceph Logo"
Image size is x=140 y=38
```

```
vagrant@mon1:~$ sudo rados -p rbd ls
image_test
```

```
vagrant@mon1:~$ sudo rados -p rbd ls
image_test
vagrant@mon1:~$ sudo rados -p rbd listxattr image_test
comment
xres
yres
```

```
vagrant@mon1:~$ sudo rados -p rbd getxattr image_test comment
Ceph Logo
vagrant@mon1:~$ sudo rados -p rbd getxattr image_test xres
140
vagrant@mon1:~$ sudo rados -p rbd getxattr image_test yres
38
```

```
vagrant@mon1:~$ sudo ./atomic
Connected to the rados cluster
Connected to pool: rbd
Object: atomic_object has been written to transaction
Would you like to abort transaction? (Y/N)? y
Transaction has been aborted, so object will not actually be written
vagrant@mon1:~$ sudo rados -p rbd ls
```

```
vagrant@mon1:~$ sudo ./atomic
Connected to the rados cluster
Connected to pool: rbd
Object: atomic_object has been written to transaction
Would you like to abort transaction? (Y/N)? n
Attribute has been written to transaction
We wrote the transaction containing our object and attributeatomic_object
vagrant@mon1:~$ sudo rados -p rbd ls
atomic_object
vagrant@mon1:~$ sudo rados -p rbd getxattr atomic_object atomic_attribute
I am an atomic attribute
```

```
vagrant@mon1:~$ sudo ./watcher
Created the rados object.
Read the config file.

Connected to the cluster.
Creating Watcher on object rbd/my_object
```

```
vagrant@mon1:~$ sudo rados -p rbd notify my_object "Hello There!"
reply client.24135 cookie 29079312 : 8 bytes
00000000  52 65 63 69 65 76 65 64                           |Recieved|
00000008
```

```
vagrant@mon1:~$ sudo ./watcher
Created the rados object.
Read the config file.

Connected to the cluster.
Creating Watcher on object rbd/my_object
Message from Notifier: Hello There!
```

Chapter 6: Distributed Computation with Ceph RADOS Classes

```
vagrant@mon1:~$ sudo python rados_lua.py LowerObject
THIS STRING WAS IN LOWERCASE
```

```
vagrant@ansible:~$ git clone https://github.com/ceph/ceph.git
Cloning into 'ceph'...
remote: Counting objects: 500133, done.
remote: Compressing objects: 100% (21/21), done.
remote: Total 500133 (delta 12), reused 2 (delta 2), pack-reused 500110
Receiving objects: 100% (500133/500133), 203.37 MiB | 2.38 MiB/s, done.
Resolving deltas: 100% (394234/394234), done.
Checking connectivity... done.
```

```
-- Configuring done
-- Generating done
-- Build files have been written to: /home/vagrant/ceph/build
+ cat
+ echo 40000
+ echo done.
done.
```

```
vagrant@ansible:~/ceph/build$ make cls_md5
Scanning dependencies of target cls_md5
[  0%] Building CXX object src/cls/CMakeFiles/cls_md5.dir/md5/cls_md5.cc.o
[100%] Linking CXX shared library ../../lib/libcls_md5.so
[100%] Built target cls_md5
```

```
vagrant@osd2:~$ sudo scp vagrant@ansible:/home/vagrant/ceph/build/lib/libcls_md5.so* /usr/lib/rados-classes/
vagrant@ansible's password:
libcls_md5.so                                                 100%  155KB 155.0KB/s   00:00
libcls_md5.so.1                                               100%  155KB 155.0KB/s   00:00
libcls_md5.so.1.0.0                                           100%  155KB 155.0KB/s   00:00
```

```
2017-05-10 19:47:57.251739 7fdb99ca2700  1 leveldb: Compacting 4@0 + 4@1 files
2017-05-10 19:47:57.260570 7fdba409fa40  1 journal _open /var/lib/ceph/osd/ceph-1/journal fd 28: 1073741824 bytes, bl
ock size 4096 bytes, directio = 1, aio = 1
2017-05-10 19:47:57.280605 7fdba409fa40  1 journal _open /var/lib/ceph/osd/ceph-1/journal fd 28: 1073741824 bytes, bl
ock size 4096 bytes, directio = 1, aio = 1
2017-05-10 19:47:57.283291 7fdba409fa40  1 filestore(/var/lib/ceph/osd/ceph-1) upgrade
2017-05-10 19:47:57.300701 7fdba409fa40  0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:46: loading cls_md5
2017-05-10 19:47:57.301246 7fdba409fa40  0 <cls> /tmp/buildd/ceph-11.2.0/src/cls/cephfs/cls_cephfs.cc:198: loading ce
phfs
2017-05-10 19:47:57.308766 7fdba409fa40  0 <cls> /tmp/buildd/ceph-11.2.0/src/cls/hello/cls_hello.cc:296: loading cls_
hello
2017-05-10 19:47:57.318132 7fdba409fa40  0 osd.1 279 crush map has features 2200130813952, adjusting msgr requires fo
r clients
2017-05-10 19:47:57.318940 7fdba409fa40  0 osd.1 279 crush map has features 2200130813952 was 8705, adjusting msgr re
quires for mons
2017-05-10 19:47:57.318966 7fdba409fa40  0 osd.1 279 crush map has features 2200130813952, adjusting msgr requires fo
r osds
```

```
vagrant@mon1:~$ g++ rados_class_md5.cc -o rados_class_md5 -lrados -std=c++11
vagrant@mon1:~$ g++ rados_md5.cc -o rados_md5 -lrados -lcrypto -std=c++11
```

```
vagrant@mon1:~$ time sudo ./rados_md5
Connected to the Ceph cluster
Connected to pool: rbd

real    0m4.708s
user    0m0.084s
sys     0m1.008s
```

```
vagrant@mon1:~$ sudo rados -p rbd getxattr LowerObject MD5
9d40bae4ff2032c9eff59806298a95bdvagrant@mon1:~$
```

```
vagrant@mon1:~$ time sudo ./rados_class_md5
Connected to the Ceph cluster
Connected to pool: rbd

real    0m0.038s
user    0m0.004s
sys     0m0.012s
```

```
vagrant@mon1:~$ sudo rados -p rbd getxattr LowerObject MD5
9d40bae4ff2032c9eff59806298a95bdvagrant@mon1:~$
```

```
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:984 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:985 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:986 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:987 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:988 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:989 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:990 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:991 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:992 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:993 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:994 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:995 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:996 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:997 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:998 - 9d40bae4ff2032c9eff59806298a95bd
0 <cls> /home/vagrant/ceph/src/cls/md5/cls_md5.cc:30: Loop:999 - 9d40bae4ff2032c9eff59806298a95bd
```

Chapter 7: Monitoring Ceph

```
root@mon1:/home/vagrant# ceph osd pool create base 64 64 replicated
pool 'base' created
root@mon1:/home/vagrant# ceph osd pool create top 64 64 replicated
pool 'top' created
```

```
root@mon1:/home/vagrant# ceph osd tier add base top
pool 'top' is now (or already was) a tier of 'base'
```

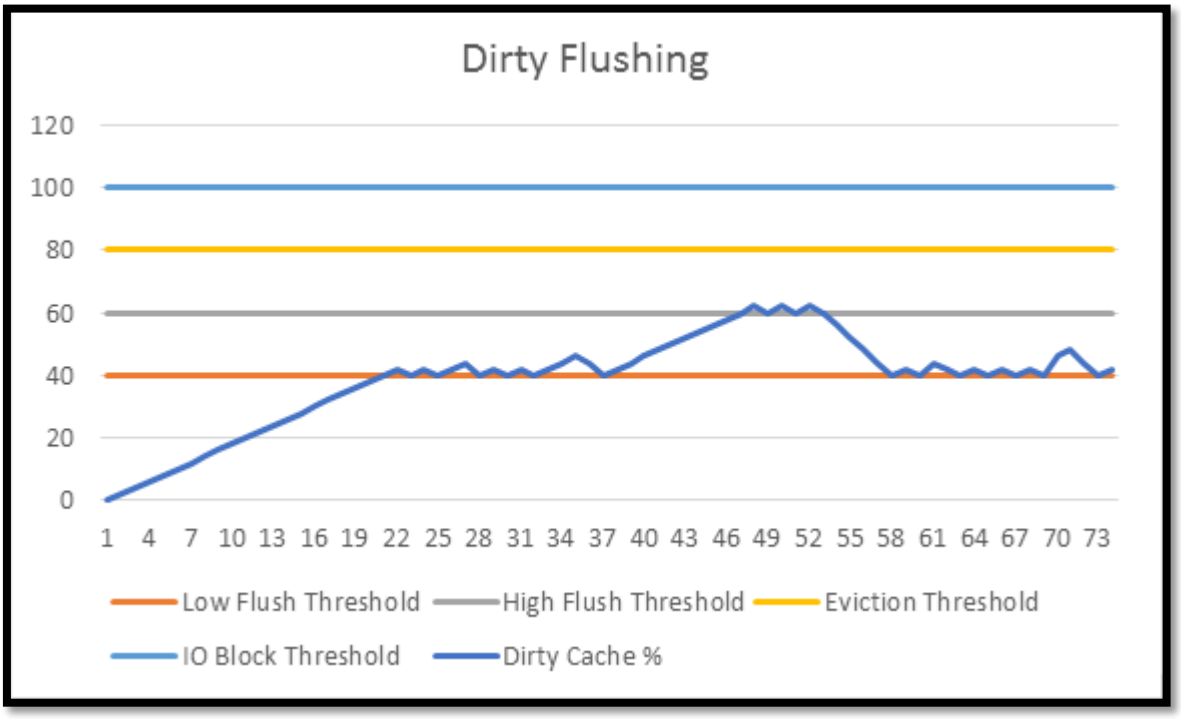
```
root@mon1:/home/vagrant# ceph osd tier cache-mode top writeback
set cache-mode for pool 'top' to writeback
```

```
root@mon1:/home/vagrant# ceph osd tier set-overlay base top
overlay for 'base' is now (or already was) 'top'
```

```
root@mon1:/home/vagrant# ceph osd pool set top hit_set_type bloom
set pool 5 hit_set_type to bloom
root@mon1:/home/vagrant# ceph osd pool set top hit_set_count 10
set pool 5 hit_set_count to 10
root@mon1:/home/vagrant# ceph osd pool set top hit_set_period 60
set pool 5 hit_set_period to 60
root@mon1:/home/vagrant# ceph osd pool set top target_max_bytes 100000000
set pool 5 target_max_bytes to 100000000
```

```
root@mon1:/home/vagrant# ceph osd pool set top cache_target_dirty_ratio 0.4
set pool 5 cache_target_dirty_ratio to 0.4
root@mon1:/home/vagrant# ceph osd pool set top cache_target_full_ratio 0.8
set pool 5 cache_target_full_ratio to 0.8
```

```
root@mon1:/home/vagrant# ceph osd pool set top cache_min_flush_age 60
set pool 5 cache_min_flush_age to 60
root@mon1:/home/vagrant# ceph osd pool set top cache_min_evict_age 60
set pool 5 cache_min_evict_age to 60
```


Dirty Flushing
120
100
80
60
40
20
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73
Low Flush Threshold
High Flush Threshold
Eviction Threshold
IO Block Threshold
Dirty Cache %

Chapter 8: Tiering with Ceph

```
vagrant@ansible:~$ sudo apt-get install graphite-api graphite-carbon graphite-web
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  fontconfig-config fonts-dejavu-core javascript-common libcairo2 libfontconfig1 libgdk-pixbuf2.0-0
  libgdk-pixbuf2.0-common libjbig0 libjpeg-turbo8 libjpeg8 libjs-jquery libjs-prototype libjs-scriptaculous
  libpixman-1-0 libtiff5 libx11-6 libx11-data libxau6 libxcb-render0 libxcb-shm0 libxcb1 libxdmcp6 libxext6
  libxrender1 python-attr python-cairo python-cffi-backend python-cryptography python-django python-django-common
  python-django-tagging python-enum34 python-idna python-ipaddress python-openssl python-pam python-pyasn1
  python-pyasn1-modules python-pyparsing python-serial python-service-identity python-simplejson python-sqlparse
  python-twisted-bin python-twisted-core python-tz python-whisper python-zope.interface python3-cairocffi python3-cffi
  python3-cffi-backend python3-cryptography python3-flask python3-idna python3-itsdangerous python3-jinja2
  python3-markupsafe python3-openssl python3-ply python3-pyasn1 python3-pycparser python3-pyinotify python3-pyparsing
  python3-structlog python3-tz python3-tzlocal python3-werkzeug python3-xcffib python3-yaml
```

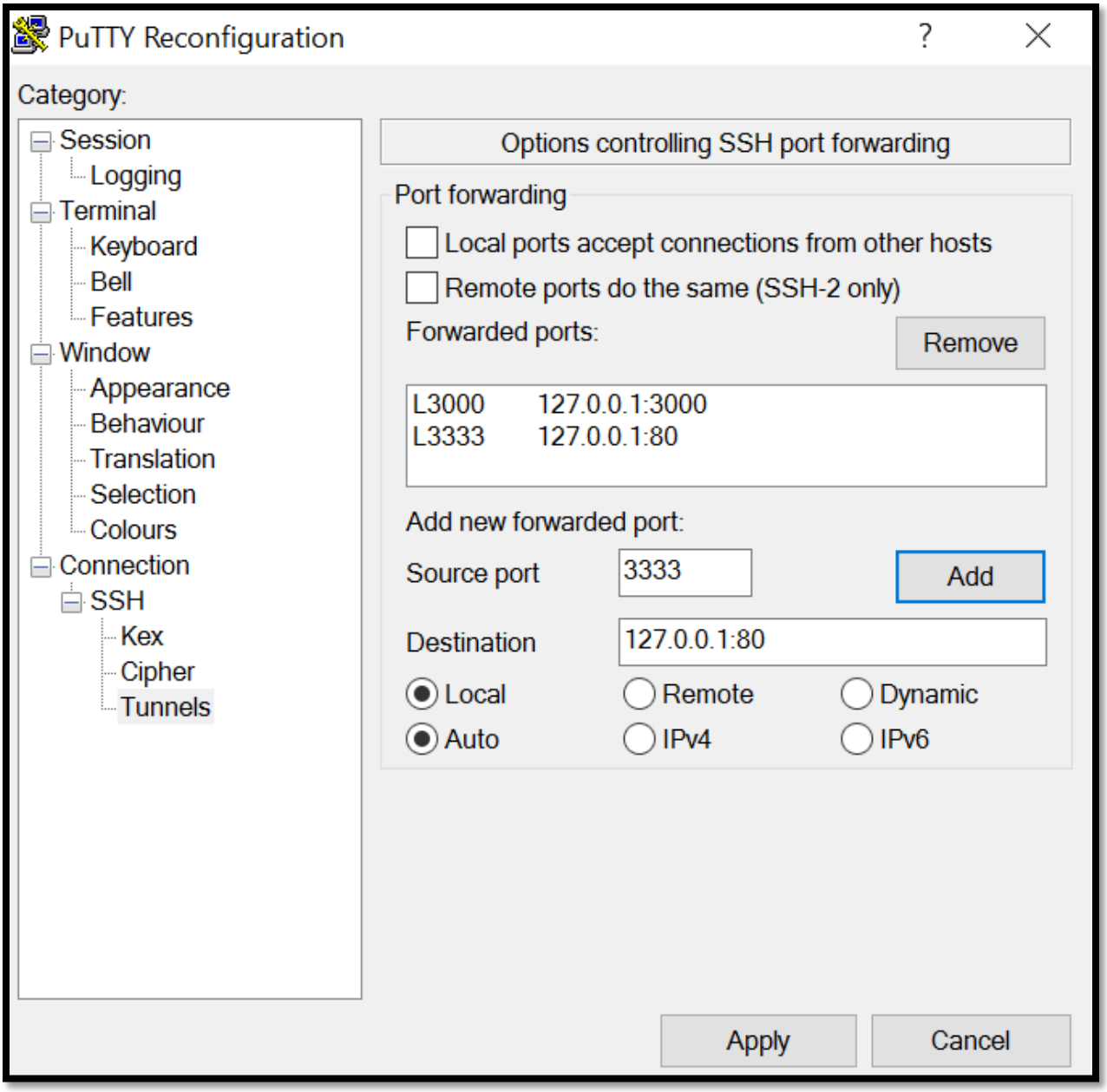
```
You have installed Django's auth system, and don't have any superusers defined.
Would you like to create one now? (yes/no): yes
Username (leave blank to use 'root'):
Email address:
Password:
Password (again):
Superuser created successfully.
```

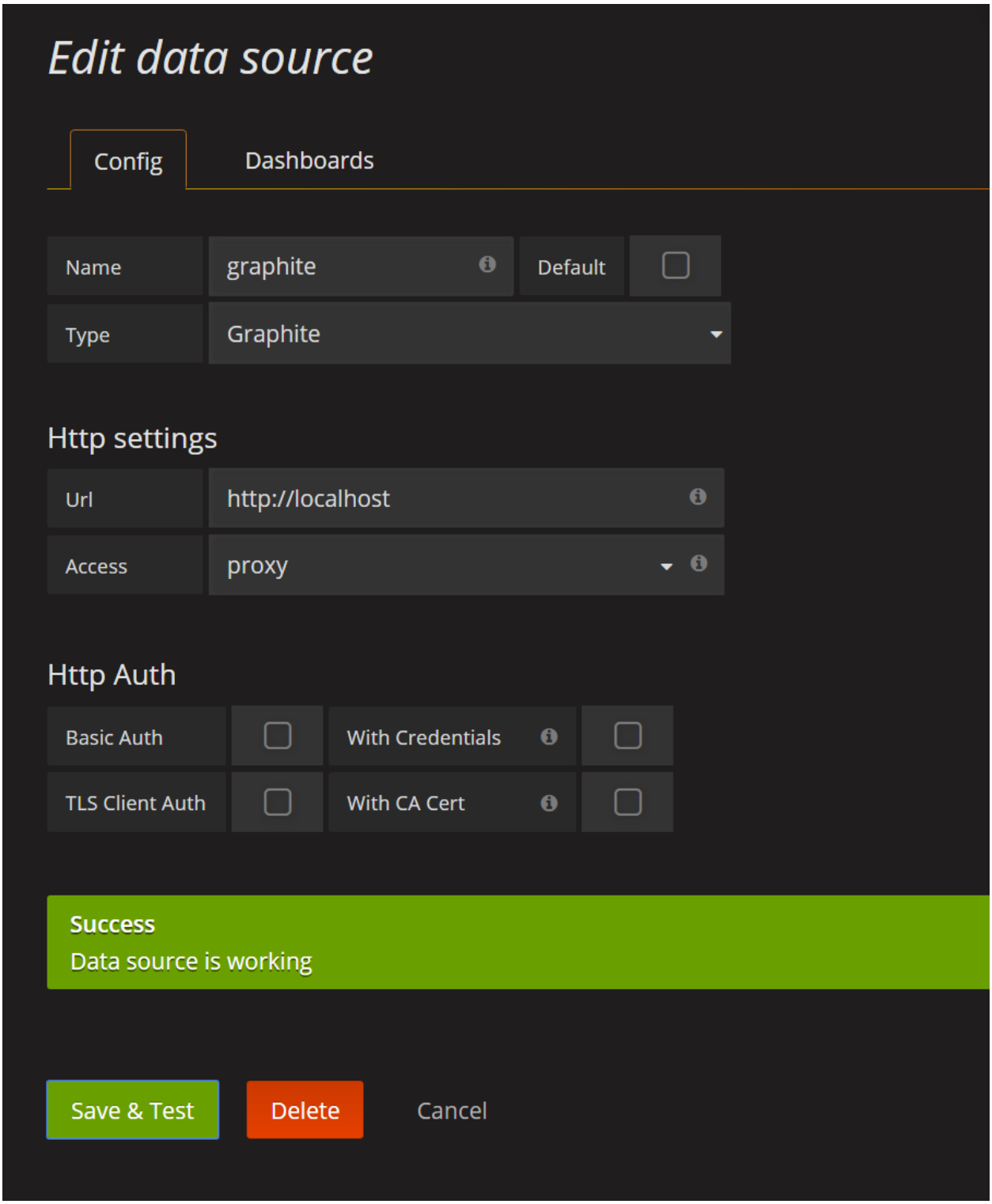
```
vagrant@ansible:~$ sudo apt-get install apache2 libapache2-mod-wsgi
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  apache2-bin apache2-data apache2-utils libaprutil1-dbd-sqlite3 libaprutil1-ldap liblua5.1-0 libpython2.7 ssl-cert
Suggested packages:
  www-browser apache2-doc apache2-suexec-pristine | apache2-suexec-custom openssl-blacklist
The following NEW packages will be installed:
  apache2 apache2-bin apache2-data apache2-utils libapache2-mod-wsgi libaprutil1-dbd-sqlite3 libaprutil1-ldap
  liblua5.1-0 libpython2.7 ssl-cert
0 upgraded, 10 newly installed, 0 to remove and 122 not upgraded.
Need to get 2,538 kB of archives.
After this operation, 9,814 kB of additional disk space will be used.
Do you want to continue? [Y/n]
```

```
vagrant@ansible:~$ sudo a2dissite 000-default
Site 000-default disabled.
To activate the new configuration, you need to run:
  service apache2 reload
vagrant@ansible:~$
```

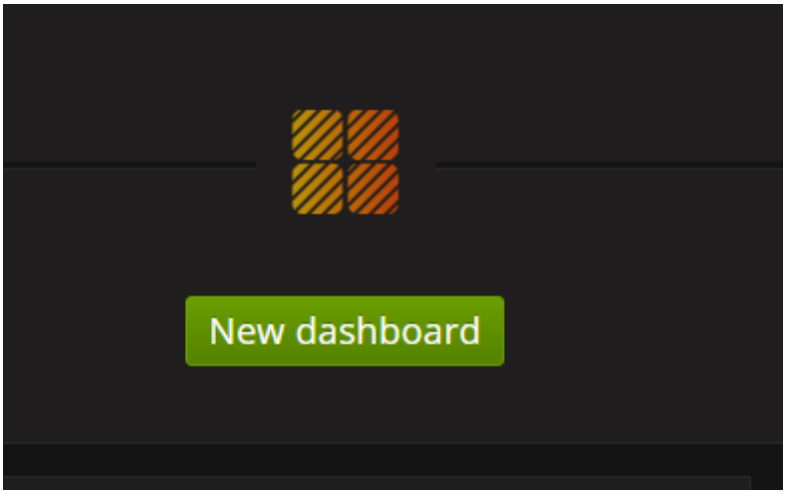
```
vagrant@ansible:~$ sudo a2ensite apache2-graphite
Enabling site apache2-graphite.
To activate the new configuration, you need to run:
  service apache2 reload
vagrant@ansible:~$
```

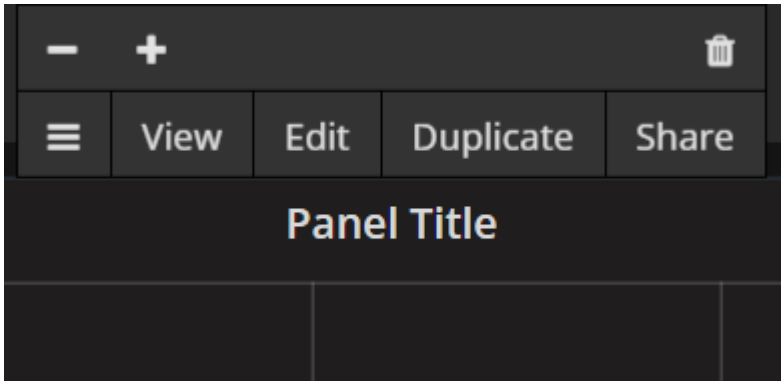
```
vagrant@ansible:~$ sudo apt-get install grafana
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  build-essential dpkg-dev fonts-font-awesome g++ g++-5 golang-1.6-go golang-1.6-race-detector-runtime golang-1.6-src
  golang-go golang-race-detector-runtime golang-src grafana-data libalgorithm-diff-perl libalgorithm-diff-xs-perl
  libalgorithm-merge-perl libjs-angularjs libjs-jquery-metadata libjs-jquery-tablesorter libjs-twitter-bootstrap
  libstdc++-5-dev pkg-config
Suggested packages:
  debian-keyring g++-multilib g++-5-multilib gcc-5-doc libstdc++6-5-dbg bzr mercurial libjs-bootstrap libstdc++-5-doc
The following NEW packages will be installed:
  build-essential dpkg-dev fonts-font-awesome g++ g++-5 golang-1.6-go golang-1.6-race-detector-runtime golang-1.6-src
  golang-go golang-race-detector-runtime golang-src grafana grafana-data libalgorithm-diff-perl
  libalgorithm-diff-xs-perl libalgorithm-merge-perl libjs-angularjs libjs-jquery-metadata libjs-jquery-tablesorter
  libjs-twitter-bootstrap libstdc++-5-dev pkg-config
0 upgraded, 22 newly installed, 0 to remove and 122 not upgraded.
Need to get 43.1 MB of archives.
After this operation, 268 MB of additional disk space will be used.
Do you want to continue? [Y/n]
```

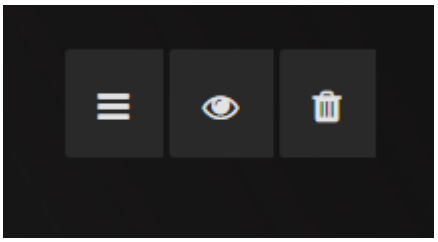


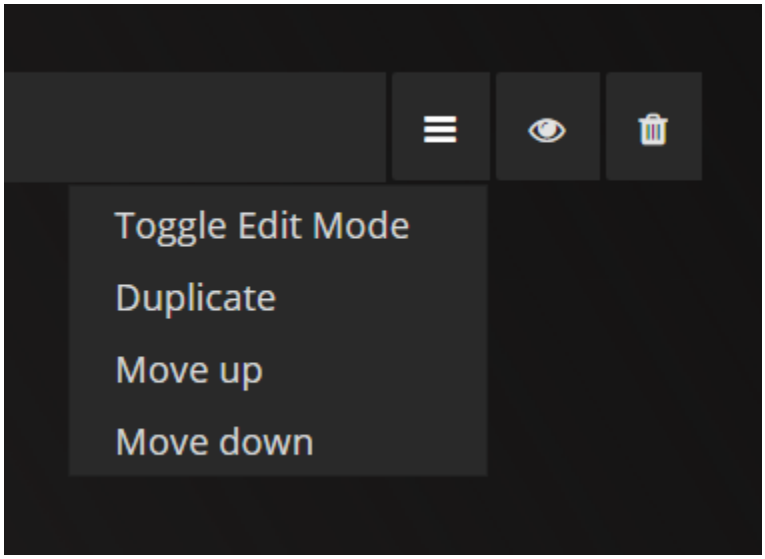

Edit data source
Config
Dashboards
Name
graphite
Default
Type
Graphite
Http settings
Url
http://localhost
Access
proxy
Http Auth
Basic Auth
With Credentials
TLS Client Auth
With CA Cert
Success
Data source is working
Save & Test
Delete
Cancel

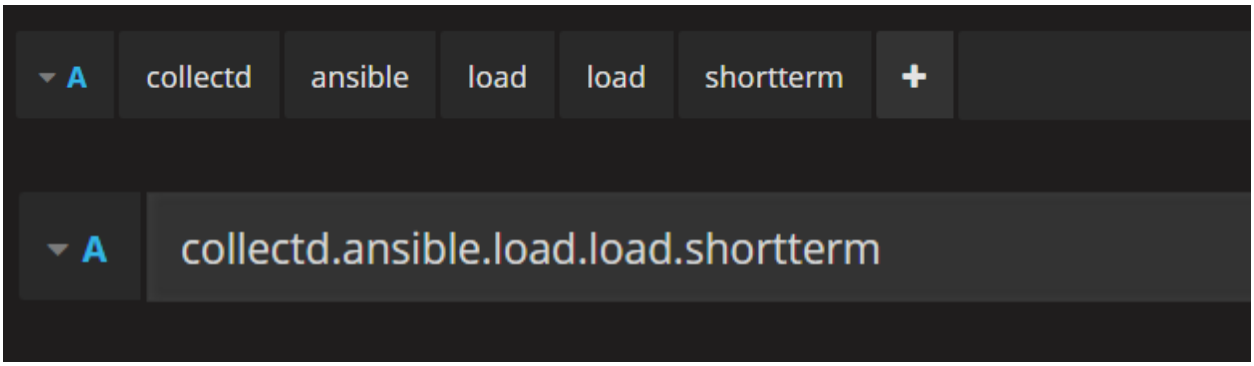
```
vagrant@ansible:~$ sudo apt-get install collectd-core
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  fontconfig libdatrie1 libdbi1 libgraphite2-3 libharfbuzz0b libltdl7 libpango-1.0-0 libpangocairo-1.0-0
  libpangoft2-1.0-0 librrd4 libthai-data libthai0 rrdtool
Suggested packages:
  collectd-dev librrds-perl liburi-perl libhtml-parser-perl libregexp-common-perl libconfig-general-perl
  apcupsd bind9 ceph hddtemp ipvsadm lm-sensors mbmon memcached mysql-server | virtual-mysql-server nginx
  notification-daemon nut openvpn olsrd pdns-server postgresql redis-server slapd time-daemon varnish
  zookeeper libatasmart4 libesmtp6 libganglia1 libhiredis0.13 libmemcached11 libmodbus5 libmysqlclient20
  libnotify4 libopenipmi0 liboping0 libowcapi-3.1-1 libpq5 libprotobuf-c1 librabbitmq4 librdkafka1
  libsensors4 libsigrok2 libsnmp30 libtokyotyrant3 libupsclient4 libvarnishapi1 libvirt0 libyajl2
  default-jre-headless
The following NEW packages will be installed:
  collectd-core fontconfig libdatrie1 libdbi1 libgraphite2-3 libharfbuzz0b libltdl7 libpango-1.0-0
  libpangocairo-1.0-0 libpangoft2-1.0-0 librrd4 libthai-data libthai0 rrdtool
0 upgraded, 14 newly installed, 0 to remove and 122 not upgraded.
Need to get 2,193 kB of archives.
After this operation, 8,419 kB of additional disk space will be used.
Do you want to continue? [Y/n]
```







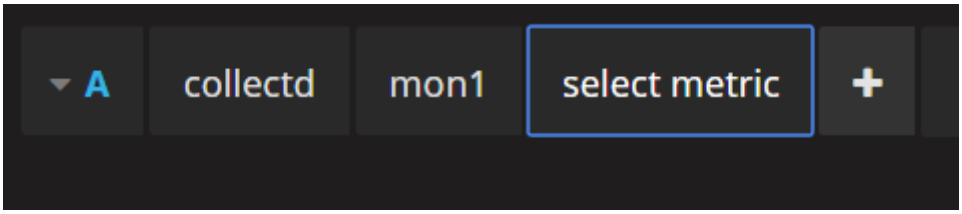


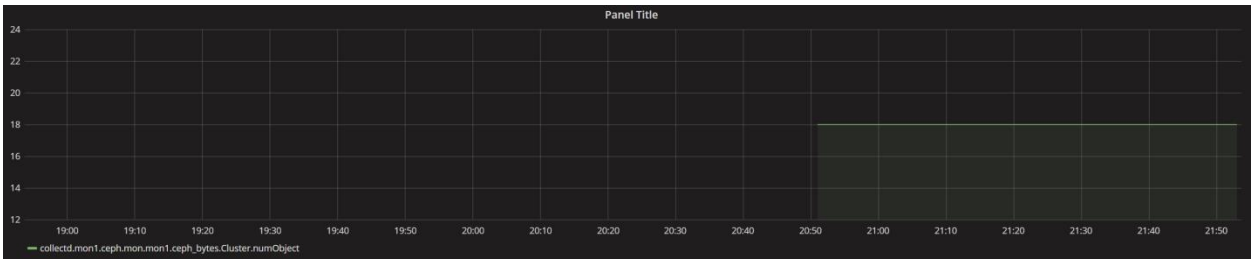



```
- hosts: mons
  gather_facts: false
  become: True
  roles:
  - ceph-mon
  - Stouts.collectd
```

```
collectd_use_ppa: yes
collectd_use_ppa_latest: yes
collectd_ppa_source: 'deb http://pkg.ci.collectd.org/deb xenial collectd-5.7'

collectd_write_graphite: yes
collectd_write_graphite_options:
  Host: "ansible"
  Port: 2003
  Prefix: collectd.
  # Postfix: .collectd
  Protocol: tcp
  AlwaysAppendDS: false
  EscapeCharacter: _
  LogSendErrors: true
  StoreRates: true
  SeparateInstances: true
  PreserveSeparator: true

collectd_ceph: yes
```

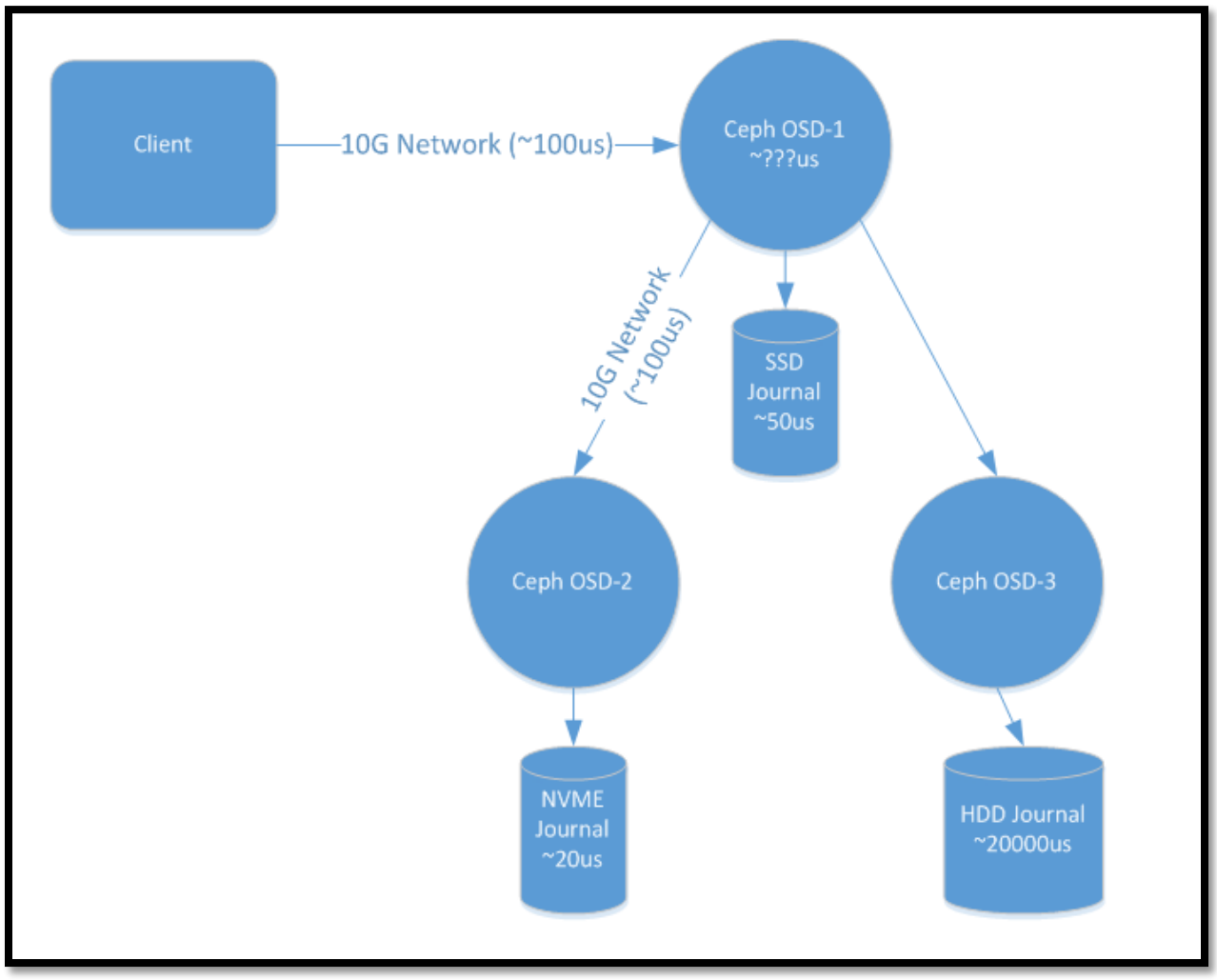
```
RUNNING HANDLER [Stouts.collectd : collectd restart] ***************************
changed: [osd2]
changed: [osd3]
changed: [osd1]

PLAY RECAP *********************************************************************
mon1                       : ok=67   changed=8    unreachable=0    failed=0
mon2                       : ok=61   changed=6    unreachable=0    failed=0
mon3                       : ok=61   changed=6    unreachable=0    failed=0
osd1                       : ok=65   changed=6    unreachable=0    failed=0
osd2                       : ok=63   changed=6    unreachable=0    failed=0
osd3                       : ok=63   changed=6    unreachable=0    failed=0
```





Chapter 9: Tuning Ceph

$$IOP = \frac{1\ Second}{Latency}\ (in\ seconds)$$




```
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
```

```
------------------------------------------------------------
Client connecting to 10.1.111.1, TCP port 5001
TCP window size:  325 KByte (default)
------------------------------------------------------------
[  3] local 10.1.111.5 port 59172 connected with 10.1.111.1 port 5001
[ ID] Interval       Transfer     Bandwidth
[  3]  0.0-10.0 sec  11.1 GBytes  9.51 Gbits/sec
```

```
The following NEW packages will be installed
  fio
0 to upgrade, 1 to newly install, 0 to remove and 121 not to upgrade.
Need to get 368 kB of archives.
After this operation, 1,572 kB of additional disk space will be used.
Get:1 http://gb.archive.ubuntu.com/ubuntu xenial/universe amd64 fio amd64 2.2.10-1ubuntu1 [368 kB]
Fetched 368 kB in 5s (69.7 kB/s)
Selecting previously unselected package fio.
(Reading database ... 579847 files and directories currently installed.)
Preparing to unpack .../fio_2.2.10-1ubuntu1_amd64.deb ...
Unpacking fio (2.2.10-1ubuntu1) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up fio (2.2.10-1ubuntu1) ...
```

```
file1: (g=0): rw=read, bs=4M-4M/4M-4M/4M-4M, ioengine=libaio, iodepth=1
fio-2.2.10
Starting 1 process
Jobs: 1 (f=1): [R(1)] [100.0% done] [100.0MB/0KB/0KB /s] [25/0/0 iops] [eta 00m:00s]
file1: (groupid=0, jobs=1): err= 0: pid=26999: Sun Mar 12 22:20:33 2017
  read : io=9496.0MB, bw=162052KB/s, iops=39, runt= 60005msec
    slat (usec): min=174, max=31852, avg=244.48, stdev=649.33
    clat (msec): min=15, max=338, avg=25.03, stdev=12.58
     lat (msec): min=15, max=338, avg=25.27, stdev=12.61
    clat percentiles (msec):
     |  1.00th=[   22],  5.00th=[   22], 10.00th=[   22], 20.00th=[   23],
     | 30.00th=[   23], 40.00th=[   23], 50.00th=[   24], 60.00th=[   24],
     | 70.00th=[   24], 80.00th=[   24], 90.00th=[   25], 95.00th=[   26],
     | 99.00th=[   80], 99.50th=[  102], 99.90th=[  182], 99.95th=[  223],
     | 99.99th=[  338]
    bw (KB  /s): min=77722, max=183641, per=100.00%, avg=163013.63, stdev=21774.82
    lat (msec) : 20=0.25%, 50=97.60%, 100=1.64%, 250=0.46%, 500=0.04%
  cpu          : usr=0.04%, sys=0.97%, ctx=2382, majf=0, minf=1036
  IO depths    : 1=100.0%, 2=0.0%, 4=0.0%, 8=0.0%, 16=0.0%, 32=0.0%, >=64=0.0%
     submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     issued    : total=r=2374/w=0/d=0, short=r=0/w=0/d=0, drop=r=0/w=0/d=0
     latency   : target=0, window=0, percentile=100.00%, depth=1

Run status group 0 (all jobs):
   READ: io=9496.0MB, aggrb=162051KB/s, minb=162051KB/s, maxb=162051KB/s, mint=60005msec, maxt=60005msec
```

```
Maintaining 16 concurrent writes of 4194304 bytes to objects of size 4194304 for up to 10 seconds or 0 objects
Object prefix: benchmark_data_ms-r1-c1-osd1_25645
  sec Cur ops   started  finished  avg MB/s  cur MB/s last lat(s)  avg lat(s)
    0       0         0         0         0         0           -           0
    1      16       121       105   419.977       420   0.0683843    0.138681
    2      15       254       239   477.947       536    0.139127    0.128909
    3      16       370       354   471.939       460   0.0906016    0.131449
    4      16       491       475   474.937       484   0.0822003    0.132115
    5      16       611       595   475.935       480    0.142927    0.132402
    6      16       731       715   476.602       480    0.110139    0.131169
    7      16       859       843    481.65       512   0.0932729    0.131419
    8      15       984       969   484.435       504    0.214752    0.131099
    9      16      1115      1099    488.38       520    0.131412    0.129911
   10      15      1229      1214   485.536       460     0.13085    0.130238
Total time run:         10.120543
Total writes made:      1230
Write size:             4194304
Object size:            4194304
Bandwidth (MB/sec):     486.14
Stddev Bandwidth:       34.0562
Max bandwidth (MB/sec): 536
Min bandwidth (MB/sec): 420
Average IOPS:           121
Stddev IOPS:            8
Max IOPS:               134
Min IOPS:               105
Average Latency(s):     0.13146
Stddev Latency(s):      0.0673788
Max latency(s):         0.703673
Min latency(s):         0.0400385
Cleaning up (deleting benchmark objects)
Clean up completed and total clean up time :0.701427
```

```
Starting 1 process
rbd engine: RBD version: 0.1.10
Jobs: 1 (f=1): [W(1)] [100.0% done] [0KB/69632KB/0KB /s] [0/68/0 iops] [eta 00m:00s]
rbd_iodepth32: (groupid=0, jobs=1): err= 0: pid=5021: Sun Mar 12 22:29:13 2017
  write: io=2020.0MB, bw=68947KB/s, iops=67, runt= 30001msec
    slat (usec): min=11, max=1741, avg=36.49, stdev=39.93
    clat (msec): min=4, max=612, avg=14.81, stdev=37.84
     lat (msec): min=4, max=612, avg=14.85, stdev=37.84
    clat percentiles (msec):
     |  1.00th=[    5],  5.00th=[    5], 10.00th=[    5], 20.00th=[    5],
     | 30.00th=[    5], 40.00th=[    5], 50.00th=[    5], 60.00th=[    6],
     | 70.00th=[    6], 80.00th=[    9], 90.00th=[   30], 95.00th=[   60],
     | 99.00th=[  186], 99.50th=[  265], 99.90th=[  420], 99.95th=[  437],
     | 99.99th=[  611]
    bw (KB  /s): min= 7231, max=174080, per=100.00%, avg=71772.68, stdev=35752.42
    lat (msec) : 10=82.13%, 20=4.95%, 50=6.73%, 100=3.66%, 250=1.93%
    lat (msec) : 500=0.54%, 750=0.05%
  cpu          : usr=0.29%, sys=0.01%, ctx=2026, majf=0, minf=0
  IO depths    : 1=100.0%, 2=0.0%, 4=0.0%, 8=0.0%, 16=0.0%, 32=0.0%, >=64=0.0%
     submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     issued    : total=r=0/w=2020/d=0, short=r=0/w=0/d=0, drop=r=0/w=0/d=0
     latency   : target=0, window=0, percentile=100.00%, depth=1

Run status group 0 (all jobs):
  WRITE: io=2020.0MB, aggrb=68947KB/s, minb=68947KB/s, maxb=68947KB/s, mint=30001msec, maxt=30001msec
```

filestore_split_multiple * abs(filestore_merge_threshold)*16

Chapter 10: Troubleshooting

```
vagrant@mon1:~$ sudo rbd create test --size=1G
vagrant@mon1:~$ sudo rbd feature disable test exclusive-lock object-map fast-diff deep-flatten
vagrant@mon1:~$ sudo rbd map test
/dev/rbd0
vagrant@mon1:~$ sudo mkfs.ext4 /dev/rbd0
mke2fs 1.42.13 (17-May-2015)
Discarding device blocks: done
Creating filesystem with 262144 4k blocks and 65536 inodes
Filesystem UUID: a95d7f60-3be3-4c15-bafd-9d37559174db
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376

Allocating group tables: done
Writing inode tables: done
Creating journal (8192 blocks): done
Writing superblocks and filesystem accounting information: done
```

```
root@mon1:/home/vagrant# rados -p rbd ls
rbd_data.1e502238e1f29.0000000000000086
rbd_data.1e502238e1f29.0000000000000000
rbd_data.1e502238e1f29.0000000000000083
rbd_data.1e502238e1f29.0000000000000060
rbd_data.1e502238e1f29.0000000000000004
```

```
root@mon1:/home/vagrant# ceph osd map rbd rbd_data.1e502238e1f29.0000000000000086
osdmap e234 pool 'rbd' (0) object 'rbd_data.1e502238e1f29.0000000000000086' -> pg 0.5ee4eb42 (0.2)
> up ([1,0,2], p1) acting ([1,0,2], p1)
```

```
vagrant@osd1:~$ sudo ls -l /var/lib/ceph/osd/ceph-0/current/0.5_head/
total 4096
-rw-r--r-- 1 ceph ceph       0 Feb  7 22:07 __head_00000005__0
-rw-r--r-- 1 ceph ceph 4194304 Mar 17 21:28 rbd\udata.1e502238e1f29.0000000000000083__head_327C8305__0
```

```
root@osd1:/home/vagrant# echo blah > /var/lib/ceph/osd/ceph-0/current/0.5_head/rbd\udata.1e502238e1f29.
0000000000000083__head_327C8305__0
```

```
root@mon1:/home/vagrant# ceph pg deep-scrub 0.5
instructing pg 0.5 on osd.2 to deep-scrub
```

```
root@mon1:/home/vagrant# ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_ERR
            1 pgs inconsistent
            3 scrub errors
            too many PGs per OSD (320 > max 300)
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 158, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon2 standbys: mon3, mon1
     osdmap e234: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v3879: 320 pgs, 4 pools, 37575 kB data, 23 objects
            229 MB used, 26665 MB / 26894 MB avail
                 319 active+clean
                   1 active+clean+inconsistent
```

```
root@mon1:/home/vagrant# ceph health detail
HEALTH_ERR 1 pgs inconsistent; 3 scrub errors; too many PGs per OSD (320 > max 300); all OSDs are running kr
aken or later but the 'require_kraken_osds' osdmap flag is not set
pg 0.5 is active+clean+inconsistent, acting [2,0,1]
3 scrub errors
```

```
root@osd1:/var/lib/ceph/osd/ceph-0/current/0.5_head# md5sum rbd\\udata.1e502238e1f29.0000000000000083__head_327C8305__0
\0d599f0ec05c3bda8c3b8a68c32a1b47  rbd\\udata.1e502238e1f29.0000000000000083__head_327C8305__0
```

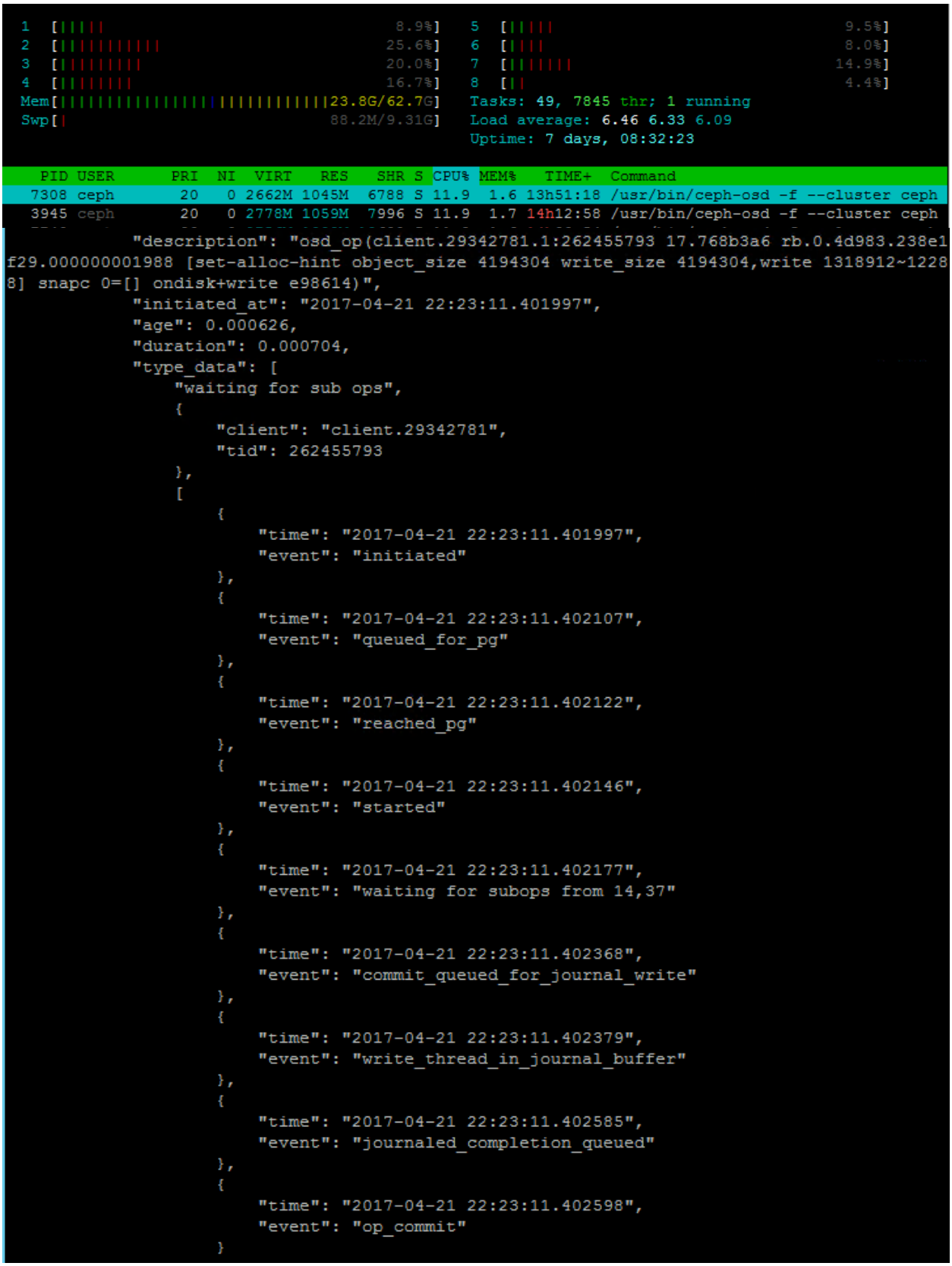
```
root@osd2:/home/vagrant# cd /var/lib/ceph/osd/ceph-2/current/0.5_head/
root@osd2:/var/lib/ceph/osd/ceph-2/current/0.5_head# md5sum rbd\\udata.1e502238e1f29.0000000000000083__
head_327C8305__0
\b5cfa9d6c8febd618f91ac2843d50a1c  rbd\\udata.1e502238e1f29.0000000000000083__head_327C8305__0
```

```
root@osd3:/home/vagrant# cd /var/lib/ceph/osd/ceph-1/current/0.5_head/
root@osd3:/var/lib/ceph/osd/ceph-1/current/0.5_head# md5sum rbd\\udata.1e502238e1f29.0000000000000083_
head_327C8305__0
\b5cfa9d6c8febd618f91ac2843d50a1c  rbd\\udata.1e502238e1f29.0000000000000083__head_327C8305__0
```

```
root@mon1:/home/vagrant# ceph pg repair 0.5
instructing pg 0.5 on osd.2 to repair
```

```
root@mon1:/home/vagrant# ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_WARN
            too many PGs per OSD (320 > max 300)
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:67
            election epoch 158, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon2 standbys: mon3, mon1
     osdmap e234: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v3900: 320 pgs, 4 pools, 37575 kB data, 23 objects
            229 MB used, 26665 MB / 26894 MB avail
                 320 active+clean
```

```
Device:         rrqm/s   wrqm/s     r/s     w/s    rkB/s    wkB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
sda               0.00     4.00    2.00   24.00    68.00  9396.00   728.00     0.34   12.92   12.00   13.00   8.77  22.80
sdc               0.00     9.00   17.00   42.00   548.00 20468.00   712.41     1.11   18.71   27.76   15.05  10.17  60.00
sdd               0.00     8.00    9.00   51.00    36.00 20888.00   697.47     1.61   26.93   27.56   26.82   8.53  51.20
sdb               0.00     2.00   10.00   18.00   416.00  8108.00   608.86     0.43   15.29   16.00   14.89   5.57  15.60
nvme0n1           0.00     0.00    0.00 1254.00     0.00 126556.00   201.84     1.66    1.32    0.00    1.32   0.06   8.00
sde               0.00     2.00   10.00    9.00   416.00  3304.00   391.58     0.26   12.84   12.40   13.33  12.42  23.60
sdf               1.00     1.00   78.00    9.00 20492.00  3820.00   558.90     1.05   12.09    9.03   38.67   4.87  42.40
sdg               0.00    18.00  117.00  108.00 29600.00 47020.00   681.07    14.59   55.38    8.17  106.52   3.70  83.20
sdh               0.00     3.00    2.00   35.00   384.00 15816.00   875.68     0.38   10.38   10.00   10.40   5.41  20.00
sdi               0.00     1.00   83.00   15.00 20508.00  7024.00   561.88     1.56   15.96    7.57   62.40   3.59  35.20
sdj               0.00     5.00   87.00   52.00 14740.00 18760.00   482.01     2.78   25.84   11.77   49.38   7.19 100.00
sdk               0.00     0.00    3.00  160.00    12.00  5748.00    70.67    11.08  350.06   17.33  356.30   1.84  30.00
sdl               0.00     0.00    6.00    0.00    24.00     0.00     8.00     0.07   11.33   11.33    0.00   9.33   5.60
```


1 [||||| 8.9%]
2 [|||||||||| 25.6%]
3 [||||||||| 20.0%]
4 [|||||||| 16.7%]
5 [||||| 9.5%]
6 [|||| 8.0%]
7 [||||||| 14.9%]
8 [|| 4.4%]
Mem[||||||||||||||||||||||23.8G/62.7G]
Swp[| 88.2M/9.31G]
Tasks: 49, 7845 thr; 1 running
Load average: 6.46 6.33 6.09
Uptime: 7 days, 08:32:23
PID USER PRI NI VIRT RES SHR S CPU% MEM% TIME+ Command
7308 ceph 20 0 2662M 1045M 6788 S 11.9 1.6 13h51:18 /usr/bin/ceph-osd -f --cluster ceph
3945 ceph 20 0 2778M 1059M 7996 S 11.9 1.7 14h12:58 /usr/bin/ceph-osd -f --cluster ceph
"description": "osd_op(client.29342781.1:262455793 17.768b3a6 rb.0.4d983.238e1f29.000000001988 [set-alloc-hint object_size 4194304 write_size 4194304,write 1318912~12288] snapc 0=[] ondisk+write e98614)",
"initiated_at": "2017-04-21 22:23:11.401997",
"age": 0.000626,
"duration": 0.000704,
"type_data": [
"waiting for sub ops",
{
"client": "client.29342781",
"tid": 262455793
},
[
{
"time": "2017-04-21 22:23:11.401997",
"event": "initiated"
},
{
"time": "2017-04-21 22:23:11.402107",
"event": "queued_for_pg"
},
{
"time": "2017-04-21 22:23:11.402122",
"event": "reached_pg"
},
{
"time": "2017-04-21 22:23:11.402146",
"event": "started"
},
{
"time": "2017-04-21 22:23:11.402177",
"event": "waiting for subops from 14,37"
},
{
"time": "2017-04-21 22:23:11.402368",
"event": "commit_queued_for_journal_write"
},
{
"time": "2017-04-21 22:23:11.402379",
"event": "write_thread_in_journal_buffer"
},
{
"time": "2017-04-21 22:23:11.402585",
"event": "journaled_completion_queued"
},
{
"time": "2017-04-21 22:23:11.402598",
"event": "op_commit"
}

```
        "probing_osds": [
            "1",
            "2"
        ],
        "blocked": "peering is blocked due to down osds",
        "down_osds_we_would_probe": [
            0
        ],
        "peering_blocked_by": [
            {
                "osd": 0,
                "current_lost_at": 0,
                "comment": "starting or marking this osd lost may let us proceed"
            }
        ]
    },
```

Chapter 11: Disaster Recovery

```
vagrant@ansible:~/ceph-ansible$ cat hosts
[mons]
site1-mon1

[osds]
site1-osd1

[ceph:children]
mons
osds
```

```
vagrant@ansible:~/ceph-ansible2$ cat hosts
[mons]
site2-mon1

[osds]
site2-osd1

[ceph:children]
mons
osds
```

```
vagrant@site1-mon1:~$ sudo ceph osd pool set rbd size 1
set pool 0 size to 1
vagrant@site1-mon1:~$ sudo ceph osd pool set rbd min_size 1
set pool 0 min_size to 1
```

```
vagrant@mon1:~$ sudo apt-get install rbd-mirror
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  rbd-mirror
0 upgraded, 1 newly installed, 0 to remove and 121 not upgraded.
Need to get 1,726 kB of archives.
After this operation, 7,240 kB of additional disk space will be used.
Get:1 http://download.ceph.com/debian-kraken xenial/main amd64 rbd-mirror amd64 11.2.0-1xenial [1,726 kB]
Fetched 1,726 kB in 1s (1,289 kB/s)
Selecting previously unselected package rbd-mirror.
(Reading database ... 54945 files and directories currently installed.)
Preparing to unpack .../rbd-mirror_11.2.0-1xenial_amd64.deb ...
Unpacking rbd-mirror (11.2.0-1xenial) ...
Processing triggers for ureadahead (0.100.0-19) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up rbd-mirror (11.2.0-1xenial) ...
ceph-rbd-mirror.target is a disabled or a static unit, not starting it.
Processing triggers for ureadahead (0.100.0-19) ...
```

```
vagrant@site1-mon1:~$ sudo rbd mirror image enable rbd/mirror_test
Mirroring enabled
```

```
vagrant@site1-mon1:~$ sudo rbd --cluster remote mirror pool status rbd --verbose
health: OK
images: 1 total
    1 replaying

mirror_test:
  global_id:   a90b307a-98ec-4835-9ea8-fc2f91b4ae37
  state:       up+replaying
  description: replaying, master_position=[object_number=3, tag_tid=1, entry_tid=2607], mirror_position=
[object_number=3, tag_tid=1, entry_tid=2607], entries_behind_master=0
  last_update: 2017-04-17 14:37:09
```

```
rbd image 'mirror_test':
        size 1024 MB in 256 objects
        order 22 (4096 kB objects)
        block_name_prefix: rbd_data.374b74b0dc51
        format: 2
        features: layering, exclusive-lock, object-map, fast-diff, deep-flatten, journaling
        flags:
        journal: 374b74b0dc51
        mirroring state: enabled
        mirroring global id: a90b307a-98ec-4835-9ea8-fc2f91b4ae37
        mirroring primary: true
```

```
rbd image 'mirror_test':
        size 1024 MB in 256 objects
        order 22 (4096 kB objects)
        block_name_prefix: rbd_data.377d2eb141f2
        format: 2
        features: layering, exclusive-lock, object-map, fast-diff, deep-flatten, journaling
        flags:
        journal: 377d2eb141f2
        mirroring state: enabled
        mirroring global id: a90b307a-98ec-4835-9ea8-fc2f91b4ae37
        mirroring primary: false
```

```
vagrant@site1-mon1:~$ sudo rbd-nbd map mirror_test
/dev/nbd0
```

```
vagrant@site1-mon1:~$ sudo mkfs.ext4 /dev/nbd0
mke2fs 1.42.13 (17-May-2015)
Discarding device blocks: done
Creating filesystem with 262144 4k blocks and 65536 inodes
Filesystem UUID: d4ff2036-a10b-4003-8a0a-144b0863b55a
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376

Allocating group tables: done
Writing inode tables: done
Creating journal (8192 blocks): done
Writing superblocks and filesystem accounting information: done
```

```
vagrant@site2-mon1:~$ sudo rbd-nbd map mirror_test
/dev/nbd0
vagrant@site2-mon1:~$ sudo mount /dev/nbd0 /mnt
vagrant@site2-mon1:~$ cat /mnt/test.txt
This is a test
```

```
vagrant@mon1:~$ git clone https://gitlab.lbader.de/kryptur/ceph-recovery.git
Cloning into 'ceph-recovery'...
remote: Counting objects: 18, done.
remote: Compressing objects: 100% (18/18), done.
remote: Total 18 (delta 6), reused 0 (delta 0)
Unpacking objects: 100% (18/18), done.
Checking connectivity... done.
```

```
vagrant@mon1:~$ sudo apt-get install sshfs
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  libboost-iostreams1.58.0 libboost-program-options1.58.0 libboost-random1.58.0 libboost-regex1.58.0 libboost-system1.58.0
  libboost-thread1.58.0 libcephfs1 libfcgi0ldbl
Use 'sudo apt autoremove' to remove them.
The following NEW packages will be installed:
  sshfs
0 upgraded, 1 newly installed, 0 to remove and 103 not upgraded.
Need to get 41.7 kB of archives.
After this operation, 138 kB of additional disk space will be used.
Get:1 http://us.archive.ubuntu.com/ubuntu xenial/universe amd64 sshfs amd64 2.5-1ubuntu1 [41.7 kB]
Fetched 41.7 kB in 0s (109 kB/s)
Selecting previously unselected package sshfs.
(Reading database ... 40714 files and directories currently installed.)
Preparing to unpack .../sshfs_2.5-1ubuntu1_amd64.deb ...
Unpacking sshfs (2.5-1ubuntu1) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up sshfs (2.5-1ubuntu1) ...
```

```
vagrant@mon1:~/ceph-recovery$ sudo ./collect_files.sh osds
Scanning ceph-0
Scanning ceph-1
Scanning ceph-2
Preparing UDATA files
UDATA files ready
Extracting VM IDs
VM IDs extracted
```

```
vagrant@mon1:~/ceph-recovery$ sudo ./assemble.sh vms/test.id 1073741824 .
1e502238e1f29
test
file_lists/1e502238e1f29.files
------------------------
CEPH RECOVERY
Assemble test with ID 1e502238e1f29
------------------------
Searching file list
file_lists/1e502238e1f29.files found
------------------------
Output Image will be ./test.raw
------------------------
There are 15 blocks found
The output file will be created as a file of size 1073741824 Bytes
The blocksize is 512
------------------------
Creating Image file...
Starting reassembly...
100% [##########################################################################################################_]
Image written to ./test.raw
```

```
vagrant@mon1:~/ceph-recovery$ sudo e2fsck test.raw
e2fsck 1.42.13 (17-May-2015)
test.raw: clean, 11/65536 files, 12635/262144 blocks
```

```
vagrant@mon1:~/ceph-recovery$ df -h
Filesystem                      Size  Used Avail Use% Mounted on
udev                            225M     0  225M   0% /dev
tmpfs                            49M  5.7M   44M  12% /run
/dev/mapper/vagrant--vg-root     38G  2.6G   34G   8% /
tmpfs                           245M     0  245M   0% /dev/shm
tmpfs                           5.0M     0  5.0M   0% /run/lock
tmpfs                           245M     0  245M   0% /sys/fs/cgroup
/dev/sda1                       472M   57M  391M  13% /boot
vagrant                         238G   95G  144G  40% /vagrant
tmpfs                            49M     0   49M   0% /run/user/1000
/dev/loop0                      976M  1.3M  908M   1% /mnt
```

```
vagrant@mon1:~/ceph-recovery$ sudo ceph osd pool set rbd min_size 1
set pool 0 min_size to 1
```

```
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_WARN
            64 pgs degraded
            26 pgs stuck unclean
            64 pgs undersized
            recovery 46/69 objects degraded (66.667%)
            too few PGs per OSD (21 < min 30)
            2/3 in osds are down
            nobackfill,norecover flag(s) set
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 258, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon1 standbys: mon2, mon3
     osdmap e398: 3 osds: 1 up, 3 in; 64 remapped pgs
            flags nobackfill,norecover,sortbitwise,require_jewel_osds
      pgmap v5286: 64 pgs, 1 pools, 37579 kB data, 23 objects
            226 MB used, 26668 MB / 26894 MB avail
            46/69 objects degraded (66.667%)
                  64 active+undersized+degraded
```

```
cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
 health HEALTH_WARN
        64 pgs degraded
        1 pgs recovering
        64 pgs stuck unclean
        64 pgs undersized
        recovery 25/69 objects degraded (36.232%)
        recovery 1/23 unfound (4.348%)
        1/3 in osds are down
        all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
 monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
        election epoch 258, quorum 0,1,2 mon1,mon2,mon3
    mgr active: mon1 standbys: mon2, mon3
 osdmap e409: 3 osds: 2 up, 3 in; 64 remapped pgs
        flags sortbitwise,require_jewel_osds
  pgmap v5319: 64 pgs, 1 pools, 37579 kB data, 23 objects
        220 MB used, 26674 MB / 26894 MB avail
        25/69 objects degraded (36.232%)
        1/23 unfound (4.348%)
              63 active+undersized+degraded
               1 active+recovering+undersized+degraded
```

```
vagrant@mon1:~$ sudo ceph health detail
HEALTH_WARN 1 pgs degraded; 1 pgs stuck unclean; recovery 2/46 objects degraded (4.348%); recovery 1/
 are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
pg 0.31 is stuck unclean for 1370.786568, current state active+degraded, last acting [2,1]
pg 0.31 is active+degraded, acting [2,1], 1 unfound
recovery 2/46 objects degraded (4.348%)
recovery 1/23 unfound (4.348%)
```

```
"recovery_state": [
    {
        "name": "Started\/Primary\/Active",
        "enter_time": "2017-03-28 21:17:56.412097",
        "might_have_unfound": [
            {
                "osd": "0",
                "status": "osd is down"
            },
            {
                "osd": "1",
                "status": "already probed"
            }
```

```
vagrant@mon1:~$ sudo ceph pg 0.31 list_missing
{
    "offset": {
        "oid": "",
        "key": "",
        "snapid": 0,
        "hash": 0,
        "max": 0,
        "pool": -9223372036854775808,
        "namespace": ""
    },
    "num_missing": 1,
    "num_unfound": 1,
    "objects": [
        {
            "oid": {
                "oid": "lost_object",
                "key": "",
                "snapid": -2,
                "hash": 1434772465,
                "max": 0,
                "pool": 0,
                "namespace": ""
            },
            "need": "398'6",
            "have": "383'5",
            "locations": []
        }
    ],
    "more": false
```

```
vagrant@mon1:~$ ls /tmp/mon-store/
kv_backend  store.db
```

```
vagrant@mon1:~$ sudo ceph-authtool /etc/ceph/ceph.client.admin.keyring --create-keyring --gen-key -n client.admin --cap mon 'allow *' --cap
 osd 'allow *' --cap mds 'allow *'
creating /etc/ceph/ceph.client.admin.keyring
```

```
vagrant@mon1:~$ sudo cat /etc/ceph/ceph.client.admin.keyring
[mon.]
        key = AQBODeBYfJFeIRAALrl1DmvSOl6983LxfCsDpA==
        caps mon = "allow *"
[client.admin]
        key = AQAzDeBYbuP+IRAA4mi1ZnbZW41v4F8taiRPHg==
        caps mds = "allow *"
        caps mon = "allow *"
        caps osd = "allow *"
```

```
vagrant@mon1:~$ sudo monmaptool /tmp/monmap --print
monmaptool: monmap file /tmp/monmap
epoch 0
fsid d9f58afd-3e62-4493-ba80-0356290b3d9f
last_changed 2017-03-29 21:14:32.762117
created 2017-03-29 21:14:32.762117
0: 192.168.0.41:6789/0 mon.noname-a
1: 192.168.0.42:6789/0 mon.noname-b
2: 192.168.0.43:6789/0 mon.noname-c
```

```
vagrant@mon1:~$ sudo monmaptool /tmp/monmap --print
monmaptool: monmap file /tmp/monmap
epoch 0
fsid d9f58afd-3e62-4493-ba80-0356290b3d9f
last_changed 2017-03-29 21:14:32.762117
created 2017-03-29 21:14:32.762117
0: 192.168.0.41:6789/0 mon.noname-a
```

```
vagrant@mon1:~$ sudo ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_WARN
            all OSDs are running kraken or later but the 'require_kraken_osds' osdmap flag is not set
     monmap e2: 1 mons at {mon1=192.168.0.41:6789/0}
            election epoch 3, quorum 0 mon1
        mgr no daemons active
     osdmap e460: 3 osds: 3 up, 3 in
            flags sortbitwise,require_jewel_osds
      pgmap v90: 64 pgs, 1 pools, 37579 kB data, 23 objects
            174 MB used, 26720 MB / 26894 MB avail
                  64 active+clean
recovery io 199 kB/s, 0 objects/s
vagrant@mon1:~$ sudo rbd ls
test
```

```
vagrant@mon1:~$ sudo ceph osd pool set rbd size 2
set pool 0 size to 2
```

```
vagrant@mon1:~$ sudo ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_ERR
            27 pgs are stuck inactive for more than 300 seconds
            64 pgs degraded
            23 pgs stale
            27 pgs stuck inactive
            27 pgs stuck unclean
            64 pgs undersized
            recovery 18/36 objects degraded (50.000%)
            too few PGs per OSD (21 < min 30)
            2/3 in osds are down
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 10, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon2 standbys: mon3, mon1
     osdmap e22: 3 osds: 1 up, 3 in; 41 remapped pgs
            flags sortbitwise,require_jewel_osds,require_kraken_osds
      pgmap v105: 64 pgs, 1 pools, 37572 kB data, 18 objects
            233 MB used, 26661 MB / 26894 MB avail
            18/36 objects degraded (50.000%)
                  41 undersized+degraded+peered
                  23 stale+undersized+degraded+peered
```

```
pg 0.21 is stale+undersized+degraded+peered, acting [2]
pg 0.22 is stale+undersized+degraded+peered, acting [2]
pg 0.23 is stale+undersized+degraded+peered, acting [2]
pg 0.24 is undersized+degraded+peered, acting [0]
pg 0.25 is undersized+degraded+peered, acting [0]
pg 0.26 is undersized+degraded+peered, acting [0]
pg 0.27 is undersized+degraded+peered, acting [0]
pg 0.28 is undersized+degraded+peered, acting [0]
pg 0.29 is undersized+degraded+peered, acting [0]
pg 0.2a is stale+undersized+degraded+peered, acting [2]
pg 0.2b is stale+undersized+degraded+peered, acting [2]
pg 0.2c is undersized+degraded+peered, acting [0]
pg 0.2d is stale+undersized+degraded+peered, acting [2]
vagrant@osd3:~$ sudo ls -l /var/lib/ceph/osd/ceph-2/current/0.2d_head
total 0
-rw-r--r-- 1 ceph ceph 0 Apr  2 20:13 __head_0000002D__0
```

```
vagrant@osd3:~$ sudo ceph-objectstore-tool --op export --pgid 0.2a --data-path /var/lib/ceph/osd/ceph-2 --file 0.2a_export
Exporting 0.2a
Read #0:54d415a2:::rbd_data.fa68238e1f29.0000000000000060:head#
Export successful
vagrant@osd3:~$ sudo ceph-objectstore-tool --op import --data-path /var/lib/ceph/osd/ceph-3 --file 0.2a_export
Importing pgid 0.2a
Write #0:54d415a2:::rbd_data.fa68238e1f29.0000000000000060:head#
Import successful
vagrant@mon1:~$ sudo ceph -s
    cluster d9f58afd-3e62-4493-ba80-0356290b3d9f
     health HEALTH_WARN
            clock skew detected on mon.mon2
            41 pgs degraded
            64 pgs stuck unclean
            41 pgs undersized
            recovery 13/36 objects degraded (36.111%)
            recovery 5/36 objects misplaced (13.889%)
            Monitor clock skew detected
     monmap e2: 3 mons at {mon1=192.168.0.41:6789/0,mon2=192.168.0.42:6789/0,mon3=192.168.0.43:6789/0}
            election epoch 10, quorum 0,1,2 mon1,mon2,mon3
        mgr active: mon2 standbys: mon3, mon1
     osdmap e48: 4 osds: 2 up, 2 in; 23 remapped pgs
            flags sortbitwise,require_jewel_osds,require_kraken_osds
      pgmap v182: 64 pgs, 1 pools, 37572 kB data, 18 objects
            1184 MB used, 16744 MB / 17929 MB avail
            13/36 objects degraded (36.111%)
            5/36 objects misplaced (13.889%)
                  41 active+undersized+degraded
                  23 active+remapped
```