GNU/Linux-Ceph

分布式文件系统-Ceph

GNU/Linux-Ceph

一、集群

1. 启动一个 ceph 进程

1) 启动 mon 进程

```
#service ceph start  mon.node1
```

2) 启动 msd 进程

```
#service ceph start mds.node1
```

3) 启动 osd 进程

```
#service ceph start osd.0
```

GNU/Linux-Ceph

一、集群

//* 以下在 ceph 客户端处执行

2. 查看机器的监控状态
   # ceph health
   HEALTH_OK

3. 查看 ceph 的实时运行状态
   # ceph -w

GNU/Linux-Ceph

一、集群

4. 检查信息状态信息
   [root@client ~]# ceph -s

5. 查看 ceph 存储空间
   [root@client ~]# ceph df

GNU/Linux-Ceph

一、集群

6. 删除一个节点的所有的 ceph 数据包

[root@node1 ~]# ceph-deploy purge node1

[root@node1 ~]# ceph-deploy purgedata node1

GNU/Linux-Ceph

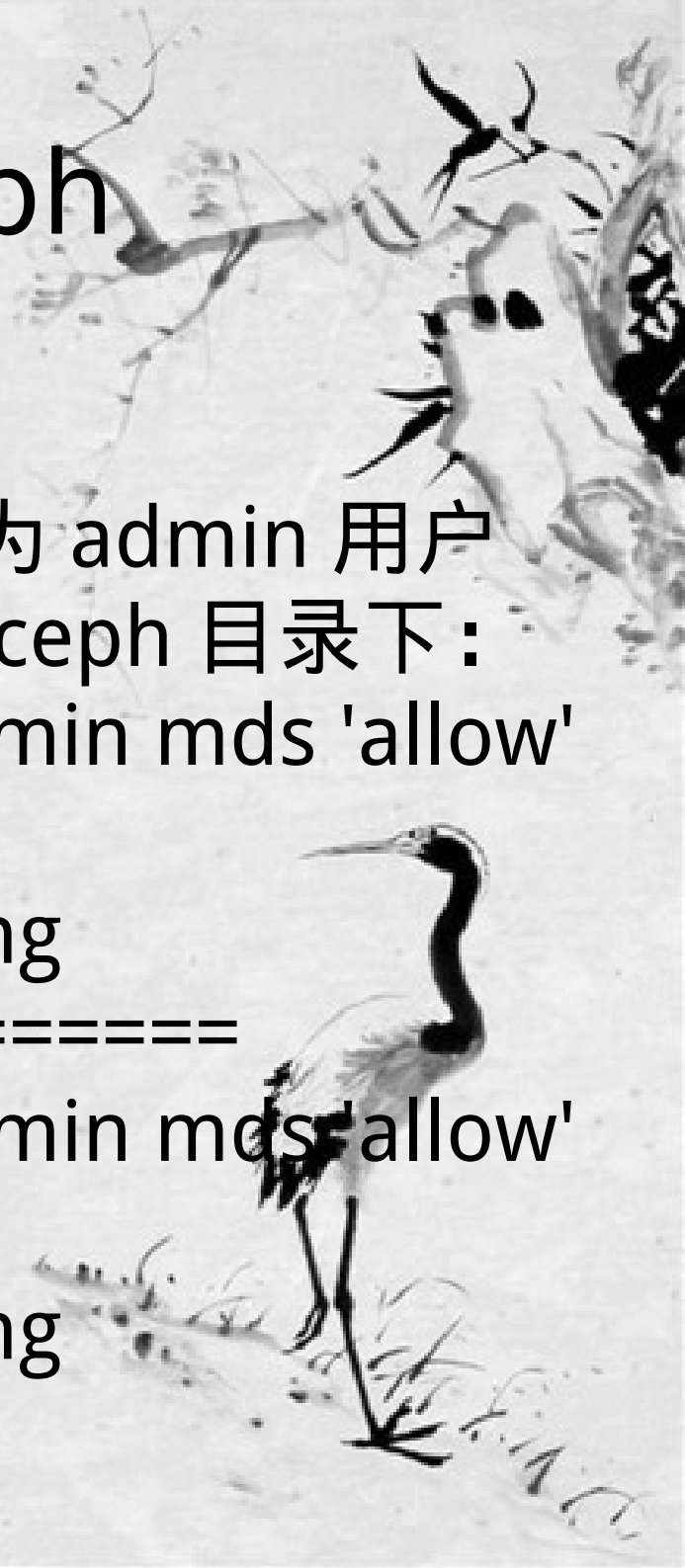
一、集群

7. 为 ceph 创建一个 admin 用户并为 admin 用户创建一个密钥，把密钥保存到 /etc/ceph 目录下：

#ceph auth get-or-create client.admin mds 'allow' osd 'allow *' mon 'allow *' > /etc/ceph/ceph.client.admin.keyring

或 ==========

#ceph auth get-or-create client.admin mds 'allow' osd 'allow *' mon 'allow *' -o /etc/ceph/ceph.client.admin.keyring



GNU/Linux-Ceph

一、集群

8. 为 osd.0 创建一个用户并创建一个 key
#ceph auth get-or-create osd.0 mon 'allow rwx' osd 'allow *' -o /var/lib/ceph/osd/ceph-0/keyring

9. 为 mds.node1 创建一个用户并创建一个 key
#ceph auth get-or-create mds.node1 mon 'allow rwx' osd 'allow *' mds 'allow *' -o /var/lib/ceph/mds/ceph-node1/keyring

GNU/Linux-Ceph

一、集群

10. 查看 ceph 集群中的认证用户及相关的 key
#ceph auth list

11. 删除集群中的一个认证用户
#ceph auth del osd.0

12. 查看集群的详细配置
[root@node1 ~]# ceph daemon mon.node1 config show | more

GNU/Linux-Ceph

一、集群

13. 查看集群健康状态细节

[root@admin ~]# ceph health detail

14. 查看 ceph log 日志所在的目录
[root@node1 ~]# ceph-conf --name mon.node1 --show-config-value log_file

GNU/Linux-Ceph

二、 mon

1. 查看 mon 的状态信息
[root@client ~]# ceph mon stat

2. 查看 mon 的选举状态
[root@client ~]# ceph quorum_status

3. 查看 mon 的映射信息
[root@client ~]# ceph mon dump

GNU/Linux-Ceph

二、 mon

4. 删除一个 mon 节点
[root@node1 ~]# ceph mon remove node1

5. 获得一个正在运行的 mon map，并保存在 1.txt 文件中
[root@node3 ~]# ceph mon getmap -o 1.txt

6. 查看上面获得的 map
[root@node3 ~]# monmaptool --print 1.txt

二、 mon

7. 把上面的 mon map 注入新加入的节点
#ceph-mon -i node4 --inject-monmap 1.txt

8. 查看 mon 的 amin socket
[root@node1 ~]# ceph-conf --name mon.node1 --show-config-value admin_socket

9. 查看 mon 的详细状态
[root@node1 ~]# ceph daemon mon.node1 mon_status

GNU/Linux-Ceph

二、 mon

10. 删除一个 mon 节点

```
[root@os-node1 ~]# ceph mon remove os-node1
```

GNU/Linux-Ceph

三、 msd

1. 查看 msd 状态
[root@client ~]# ceph mds stat

2. 查看 msd 的映射信息
[root@client ~]# ceph mds dump

3. 删除一个 mds 节点
[root@node1 ~]# ceph mds rm 0 mds.node1

GNU/Linux-Ceph

三、 OSD

1. 查看 ceph osd 运行状态
[root@client ~]# ceph osd stat

2. 查看 osd 映射信息
[root@client ~]# ceph osd dump

3. 查看 osd 的目录树
[root@client ~]# ceph osd tree

GNU/Linux-Ceph

三、 OSD

4.down 掉一个 osd 硬盘 (down 掉 osd.0 节点)
[root@node1 ~]# ceph osd down 0

5. 在集群中删除一个 osd 硬盘
[root@node4 ~]# ceph osd rm 0

6. 在集群中删除一个 osd 硬盘 crush map
[root@node1 ~]# ceph osd crush rm osd.0

GNU/Linux-Ceph

三、 OSD

7. 在集群中删除一个 osd 的 host 节点
[root@node1 ~]# ceph osd crush rm node1

8. 查看最大 osd 的个数
[root@node1 ~]# ceph osd getmaxosd

9. 设置最大的 osd 的个数（当扩大 osd 节点的时候必须扩大这个值）
[root@node1 ~]# ceph osd setmaxosd 10

三、OSD

10. 设置 osd crush 的权重为 1.0
#ceph osd crush set {id} {weight} [{loc1} [{loc2} …]]

例 :
[root@admin ~]# ceph osd crush set 3 3.0 host=node4
set item id 3 name 'osd.3' weight 3 at location {host=node4} to crush map
[root@admin ~]# ceph osd tree

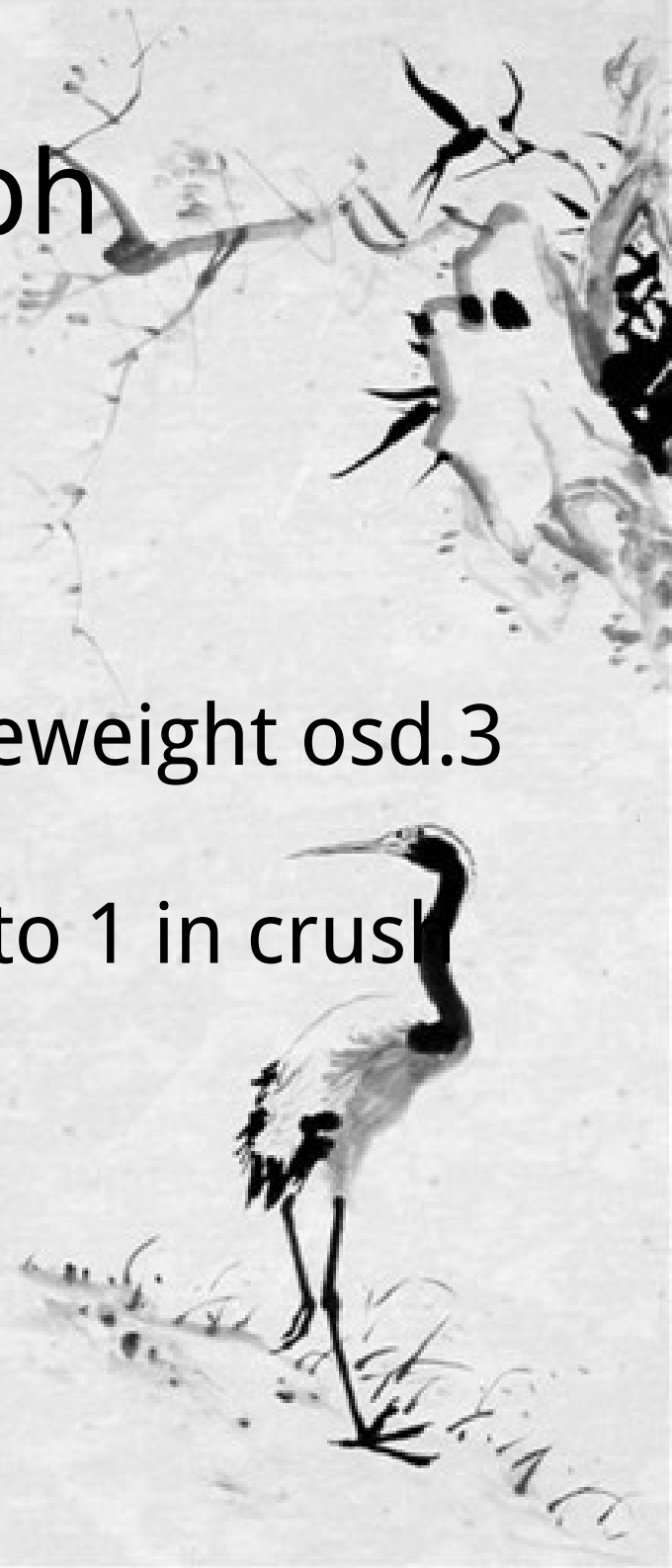
GNU/Linux-Ceph

三、 OSD

或者用下面的方式

```
[root@admin ~]# ceph osd crush reweight osd.3 1.0
reweighted item id 3 name 'osd.3' to 1 in crush map

[root@admin ~]# ceph osd tree
```



GNU/Linux-Ceph

三、 OSD

12. 把一个 osd 节点逐出集群
[root@admin ~]# ceph osd out osd.3
marked out osd.3.

[root@admin ~]# ceph osd tree

GNU/Linux-Ceph

三、 OSD

13. 把逐出的 osd 加入集群
[root@admin ~]# ceph osd in osd.3
marked in osd.3.

[root@admin ~]# ceph osd tree

GNU/Linux-Ceph

三、 OSD

14. 暂停 osd （暂停后整个集群不再接收数据）

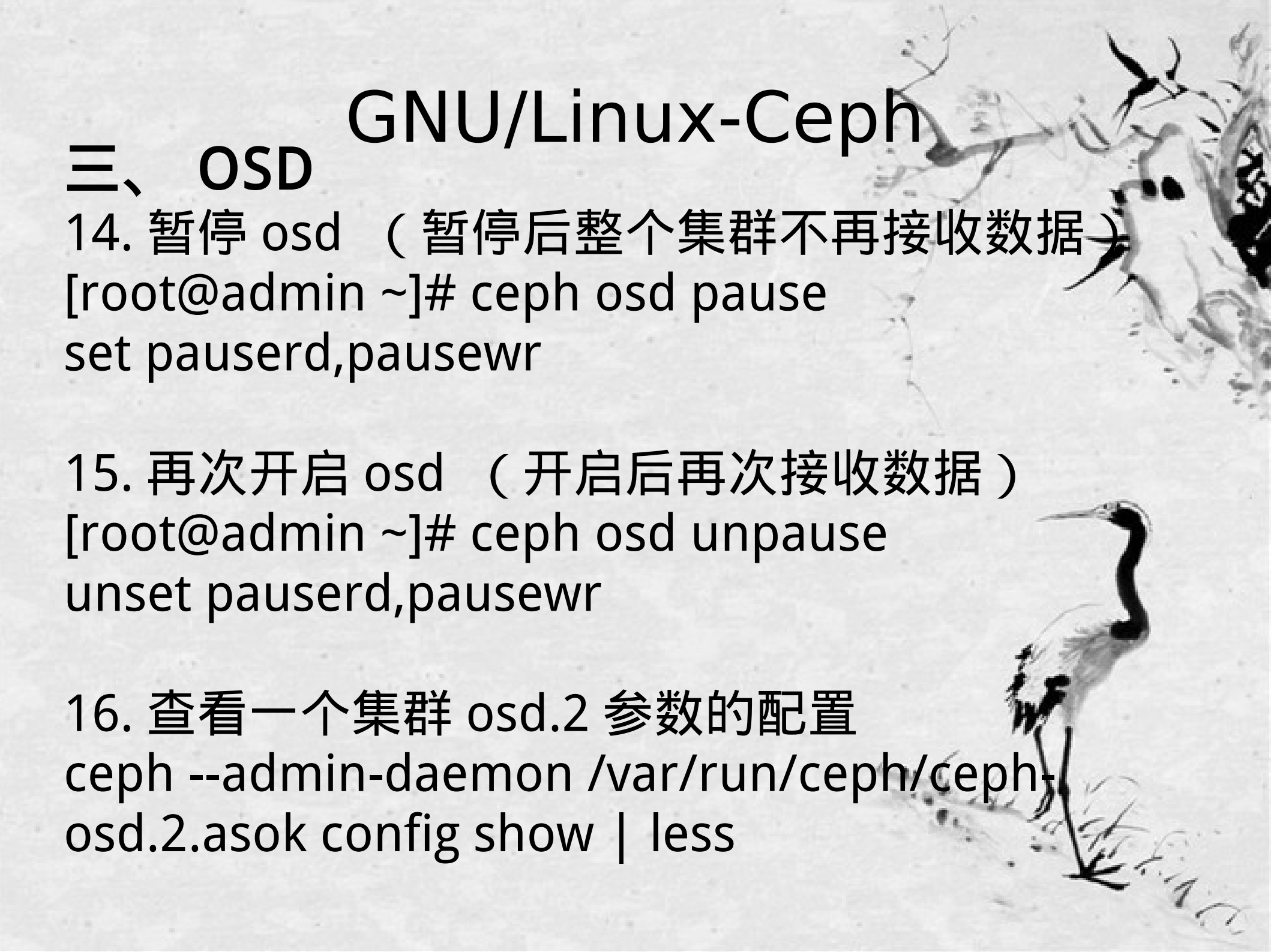
```
[root@admin ~]# ceph osd pause
set pauserd,pausewr
```

15. 再次开启 osd （开启后再次接收数据）

```
[root@admin ~]# ceph osd unpause
unset pauserd,pausewr
```

16. 查看一个集群 osd.2 参数的配置

```
ceph --admin-daemon /var/run/ceph/ceph-osd.2.asok config show | less
```



GNU/Linux-Ceph

四、 PG 组

1. 查看 pg 组的映射信息
[root@client ~]# ceph pg dump

2. 查看一个 PG 的 map
[root@client ~]# ceph pg map 0.3f
osdmap e88 pg 0.3f (0.3f) -> up [0,2] acting [0,2]
其中的 [0,2] 代表存储在 osd.0 、 osd.2 节点， osd.0 代表主副本的存储位置

GNU/Linux-Ceph

四、 PG 组

3. 查看 PG 状态
[root@client ~]# ceph pg stat

4. 查询一个 pg 的详细信息
[root@client ~]# ceph pg 0.26 query

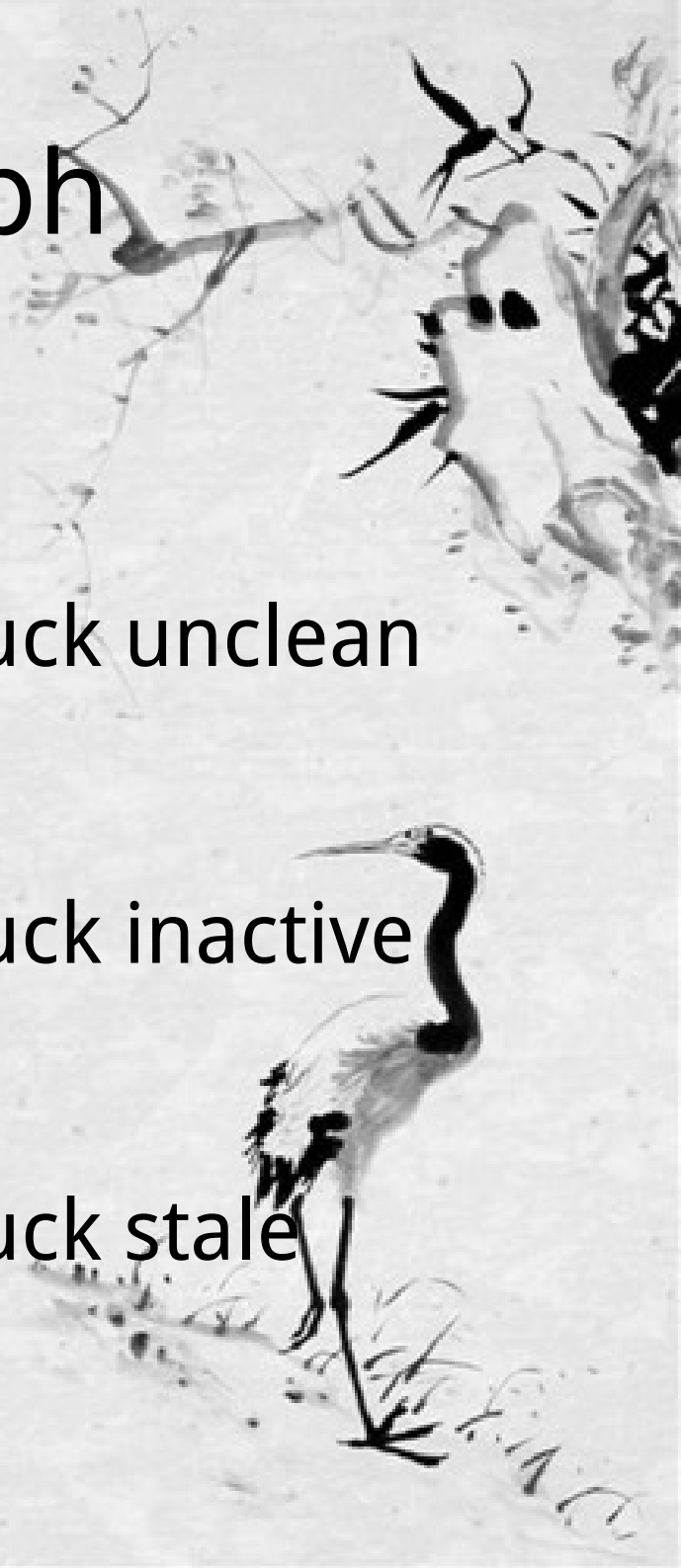
四、 PG 组

5. 查看 pg 中 stuck 的状态

```
[root@client ~]# ceph pg dump_stuck unclean
ok

[root@client ~]# ceph pg dump_stuck inactive
ok

[root@client ~]# ceph pg dump_stuck stale
ok
```



GNU/Linux-Ceph

四、 PG 组

6. 显示一个集群中的所有的 pg 统计
#ceph pg dump --format plain

7. 恢复一个丢失的 pg
#ceph pg {pg-id} mark_unfound_lost revert

8. 显示非正常状态的 pg
#ceph pg dump_stuck inactive|unclean|stale

GNU/Linux-Ceph

五、 pool

1. 查看 ceph 集群中的 pool 数量
#[root@admin ~]# ceph osd lspools

2. 在 ceph 集群中创建一个 pool(100 为 PG 组)
#ceph osd pool create jiayuan 100

3. 为一个 ceph pool 配置配额
#ceph osd pool set-quota data max_objects 10000

五、 pool

4. 在集群中删除一个 pool
#ceph osd pool delete jiayuan jiayuan --yes-i-really-really-mean-it # 集群名字需要重复两次

5. 显示集群中 pool 的详细信息
#[root@admin ~]# rados df

6. 给一个 pool 创建一个快照
[root@admin ~]# ceph osd pool mksnap data date-snap

五、 pool GNU/Linux-Ceph

7. 删除 pool 的快照
[root@admin ~]# ceph osd pool rmsnap data date-snap

8. 查看 data 池的 pg 数量
[root@admin ~]# ceph osd pool get data pg_num

9. 设置 data 池的最大存储空间为 100T（默认是 1T)
[root@admin ~]# ceph osd pool set data target_max_bytes 100000000000000

五、 pool

10. 设置 data 池的副本数是 3
[root@admin ~]# ceph osd pool set data size 3

11. 设置 data 池能接受写操作的最小副本为 2
[root@admin ~]# ceph osd pool set data min_size 2

12. 查看集群中所有 pool 的副本尺寸
[root@admin mycephfs]# ceph osd dump | grep 'replicated size'

GNU/Linux-Ceph

五、 pool

13. 设置一个 pool 的 pg 数量
[root@admin ~]# ceph osd pool set data pg_num 100

14. 设置一个 pool 的 pgp 数量
[root@admin ~]# ceph osd pool set data pgp_num 100

GNU/Linux-Ceph

六、 rados 和 rbd

1.rados 命令使用方法

(1) 查看 ceph 集群中有多少个 pool （只是查看 pool)
[root@node-4 ~]# rados lspools

(2) 查看 ceph 集群中有多少个 pool, 并且每个 pool 容量及利用情况
[root@node-4 ~]# rados df

GNU/Linux-Ceph

六、 rados 和 rbd

1.rados 命令使用方法

(3) 创建一个 pool
[root@node-4 ~]#rados mkpool test

(4) 查看 ceph pool 中的 ceph object （这里的 object 是以块形式存储的）
[root@node-4 ~]# rados ls -p volumes | more

GNU/Linux-Ceph

六、 rados 和 rbd

1.rados 命令使用方法

(5) 创建一个对象 object

```
[root@admin-node ~]# rados create test-object
-p test

[root@admin-node ~]# rados -p test ls
```

六、 rados 和 rbd

1.rados 命令使用方法

(6) 删除一个对象
[root@admin-node ~]# rados rm test-object-1 -p test

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(1) 查看 ceph 中一个 pool 里的所有镜像
[root@node-4 ~]# rbd ls images

[root@node-4 ~]# rbd ls volumes

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(2) 查看 ceph pool 中一个镜像的信息

```
[root@node-4 ~]# rbd info -p images --image
74cb427c-cee9-47d0-b467-af217a67e60a
```

(3) 在 test 池中创建一个命名为 zhanguo 的 10000M 的镜像

```
[root@node-4 ~]# rbd create -p test --size 10000
niliu
[root@node-4 ~]# rbd -p test info niliu
```

六、 rados 和 rbd

2.rbd 命令的用法

(4) 删除一个镜像

```
[root@node-4 ~]# rbd rm  -p test  niliuimages
```

(5) 调整一个镜像的尺寸

```
[root@node-4 ~]# rbd resize -p test --size 20000
niliuimages
[root@node-4 ~]# rbd -p test info niliuimages
```

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(6) 给一个镜像创建一个快照
[root@node-4 ~]# rbd snap create
test/niliu@niliu123 ← 池 / 镜像 @ 快照

[root@node-4 ~]# rbd   snap ls -p test niliu

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(6) 给一个镜像创建一个快照
[root@node-4 ~]# rbd snap create
test/niliu@niliu123 ← 池 / 镜像 @ 快照

[root@node-4 ~]# rbd   snap ls -p test niliu

[root@node-4 ~]# rbd info test/niliu@niliu123

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(7) 查看一个镜像文件的快照

```
[root@os-node101 ~]# rbd snap ls  -p volumes
volume-7687988d-16ef-4814-8a2c-3fbd85e928e4
```

(8) 删除一个镜像文件的一个快照快照 (快照所在的池 / 快照所在的镜像文件 @ 快照)

```
[root@os-node101 ~]# rbd snap rm
volumes/volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4@snapshot-ee7862aa-825e-4004-
9587-879d60430a12
```

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(9) 删除写保护后再进行删除。

```
[root@os-node101 ~]# rbd snap unprotect
volumes/volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4@snapshot-ee7862aa-825e-4004-
9587-879d60430a12

[root@os-node101 ~]# rbd snap rm
volumes/volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4@snapshot-ee7862aa-825e-4004-
9587-879d60430a12
```

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(10) 除一个镜像文件的所有快照

```
[root@os-node101 ~]# rbd snap purge  -p
volumes volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4
```

(11) 把 ceph pool 中的一个镜像导出

```
[root@node-4 ~]# rbd export -p images --image
74cb427c-cee9-47d0-b467-af217a67e60a
/root/aaa.img
```

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(12) 导出云硬盘
[root@node-4 ~]# rbd export -p volumes --image volume-470fee37-b950-4eef-a595-d7def334a5d6 /var/lib/glance/ceph-pool/volumes/Message-JiaoBenJi-10.40.212.24

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(13) 把一个镜像导入 ceph 中（但是直接导入是不能用的，因为没有经过 openstack,openstack 是看不到的）

[root@node-4 ~]# rbd import /root/aaa.img -p images --image 74cb427c-cee9-47d0-b467-af217a67e60a

GNU/Linux-Ceph

六、 rados 和 rbd

2.rbd 命令的用法

(9) 删除写保护后再进行删除。

```
[root@os-node101 ~]# rbd snap unprotect
volumes/volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4@snapshot-ee7862aa-825e-4004-
9587-879d60430a12

[root@os-node101 ~]# rbd snap rm
volumes/volume-7687988d-16ef-4814-8a2c-
3fbd85e928e4@snapshot-ee7862aa-825e-4004-
9587-879d60430a12
```