实验环境手工配置新增一台osd主机

原有ceph集群情况

物理拓扑

root	default		
角色	mon&osd	mon&osd	mon&osd
rack	rack-01	rack-02	rack-03
host	node6-1	node6-2	noce6-3
osd进程实例名称	osd.0	osd.3	osd.6
	osd.1	osd.4	osd.7
	osd.2	osd.5	osd.8

逻辑拓扑

failure domain	sata-01		
replica domain	replica-01		
osd domain	osd-01	osd-02	osd-03
osd instance	osd.0	osd.3	osd.6
	osd.1	osd.4	osd.7
	osd.2	osd.5	osd.8

osd tree情况

```
[root@node6-3 ~]# ceph osd tree
ID  WEIGHT  TYPE NAME                     UP/DOWN REWEIGHT PRIMARY-AFFINITY
-12 0.21599 failure-domain sata-01
 -8 0.21599     replica-domain replica-01
 -9 0.07199         osd-domain osd-01
  0 0.02399             osd.0                  up  1.00000          1.00000
  1 0.02399             osd.1                  up  1.00000          1.00000
  2 0.02399             osd.2                  up  1.00000          1.00000
-10 0.07199         osd-domain osd-02
  3 0.02399             osd.3                  up  1.00000          1.00000
  4 0.02399             osd.4                  up  1.00000          1.00000
  5 0.02399             osd.5                  up  1.00000          1.00000
-11 0.07199         osd-domain osd-03
  6 0.02399             osd.6                  up  1.00000          1.00000
  7 0.02399             osd.7                  up  1.00000          1.00000
  8 0.02399             osd.8                  up  1.00000          1.00000
 -1 0.21899 root default
 -5 0.07300     rack rack-01
 -2 0.07300         host node6-1
  0 0.02399             osd.0                  up  1.00000          1.00000
  1 0.02399             osd.1                  up  1.00000          1.00000
  2 0.02399             osd.2                  up  1.00000          1.00000
 -6 0.07300     rack rack-02
 -3 0.07300         host node6-2
  3 0.02399             osd.3                  up  1.00000          1.00000
  4 0.02399             osd.4                  up  1.00000          1.00000
  5 0.02399             osd.5                  up  1.00000          1.00000
 -7 0.07300     rack rack-03
 -4 0.07300         host node6-3
  6 0.02399             osd.6                  up  1.00000          1.00000
  7 0.02399             osd.7                  up  1.00000          1.00000
  8 0.02399             osd.8                  up  1.00000          1.00000
```

实现结果

在ceph集群中新加入一台osd主机node6-4，该主机中运行3个osd进程实例

修改crush map：在物理拓扑上将3个进程实例运行单独物理主机node6-4上；在逻辑拓扑上将3个进程运行在单独的failure domain：sata-02和单独的replica domain：replica-02上

物理拓扑

root	default			
角色	mon&osd	mon&osd	mon&osd	osd
rack	rack-01	rack-02	rack-03	rack-04
host	node6-1	node6-2	noce6-3	node6-4
osd进程实例名称	osd.0	osd.3	osd.6	osd.9
	osd.1	osd.4	osd.7	osd.10
	osd.2	osd.5	osd.8	osd.11

逻辑拓扑

failure domain	sata-01			sata-02
replica domain	replica-01			replica-02
osd domain	osd-01	osd-02	osd-03	osd-04
osd instance	osd.0	osd.3	osd.6	osd.9
	osd.1	osd.4	osd.7	osd.10
	osd.2	osd.5	osd.8	osd.11

部署前提

1. 编辑ceph集群中某个mon节点的本地主机解析文件/etc/hosts，在该文件中添加新加主机node6-4的解析配置：192.168.5.52 node6-2，并将配置文件/etc/hosts复制到ceph集群中的其他主机和node6-4当中

2. 在主机node6-4上安装ceph的程序包，可以参考文档:手工创建ceph集群中的相关操作过程

添加osd主机操作过程

1.复制ceph集群中mon节点主机上ceph的client认证密钥文件到node6-4主机上

```
[root@node6-4 ~]#scp node6-1:/etc/ceph/ceph.client.admin.keyring /etc/ceph/ceph.client.admin.keyring  ##
ceph的client认证密钥文件用于实现将node6-4作为ceph的client向ceph执行相应命令
```

2.在ceph集群中生成新的osd并新建osd对应目录

```
[root@node6-4 ~]#ceph osd create

9

[root@node6-4 ~]#mkdir -p /var/lib/ceph/osd/ceph-9
```

3.对指定磁盘进行分区，创建日志分区和数据分区，并对数据分区进行文件系统格式化

```
[root@node6-4 ~]#parted -a optimal -s /dev/sdd mktable gpt
[root@node6-4 ~]#parted -a optimal -s /dev/sdd mkpart ceph 0% 15GB
[root@node6-4 ~]#parted -a optimal -s /dev/sdd mkpart ceph 15GB 100%

[root@node6-4 ~]#mkfs.xfs /dev/sdd2
```

4.挂载数据分区文件系统至指定目录

```
[root@node6-4 ~]#mount -t xfs -o rw,nodev,noexec,noatime,nodiratime,attr2,discard,inode64,logbsize=256k,noquota /dev/sdd2
/var/lib/ceph/osd/ceph-9
```

5.查看数据分区所在磁盘的wwn序列号

```
[root@node6-4 ~]#ls -la /dev/disk/by-id/ | grep sdd2 | grep wwn | awk '{print $9}'|awk -F- '{print $2}'
0x5000c50087058039
```

6.编辑主机的挂载配置文件

```
[root@node6-4 ~]#vim /etc/fstab

>>

/dev/disk/by-id/wwn-0x5000c50087058039-part2 /var/lib/ceph/osd/ceph-9 xfs
rw,noexec,nodev,noatime,nodiratime,barrier=0,discard,inode64,logbsize=256k,delaylog 0 2
```

<<

7.查看数据分区进行文件系统格式化后产生的uuid

```
[root@node6-4 ~]#lsblk -f | grep sdd2|awk '{printf $3}'
```

```
[root@node6-4 ~]#8813e49b-0cd1-4861-a06a-8d00d5439281
```

8.复制node6-1主机中ceph的配置文件到node6-4中，并修改该配置文件

```
[root@node6-4 ~]#scp node6-1:/etc/ceph/ceph.conf /etc/ceph/ceph.conf
```

```
[root@node6-4 ~]#vim /ect/ceph.conf ## 编辑复制得到的ceph的配置文件，只保留其中的global配置段，并添加如下内容：
```

>>

```
[osd.9]
host = node6-4
osd_data = /var/lib/ceph/osd/ceph-9
osd_journal_size = 14336
osd_journal = /dev/disk/by-id/wwn-0x5000c50087058039-part1
```

<<

9. 初始化osd进程实例

```
[root@node6-4 ~]#ceph-osd -i 9 --mkfs --mkkey --osd-uuid `lsblk -f | grep sdd2 |awk '{print $3}'`
```

10.添加指定osd实例的认证信息

```
[root@node6-4 ~]#ceph auth add osd.0 osd 'allow *' mon 'allow rwx' -i /var/lib/ceph/osd/ceph-9/keyring
```

11.启动osd实例进程

```
[root@node6-4 ~]#service ceph start osd.9
```

12.查看osd进程是否正常启动

```
[root@node6-4 ~]#ceph -s
```

至此已经完成添加osd.9

执行上述的2~12步骤，来添加osd.10和osd.11

修改crush map

修改物理拓扑

1. 创建新的rack，并将该主机node6-4移动至该rack

```
[root@node6-4 ~]# ceph osd crush add-bucket rack-04 rack
added bucket rack-04 type rack to crush map
```

```
[root@node6-4 ~]# ceph osd crush move node6-4 rack=rack-04
```

```
moved item id -13 name 'node6-4' to location {rack=rack-04} in crush map
```

2. 向主机node6-4中添加osd进程实例

```
[root@node6-4 ~]#ceph osd cursh add osd.9 0.024 host=node6-4
```

```
[root@node6-4 ~]#ceph osd cursh add osd.10 0.024 host=node6-4

[root@node6-4 ~]#ceph osd cursh add osd.11 0.024 host=node6-4
```

3. 将rack-04移动至default中

```
[root@node6-4 ~]# ceph osd crush move rack-04 root=default
moved item id -14 name 'rack-04' to location {root=default} in crush map
```

修改逻辑拓扑

1. 新建replica domain: replica-02

```
[root@node6-4 ~]# ceph osd crush add-bucket replica-02 replica-domain

added bucket replica-02 type replica-domain to crush map
```

2. 新建osd domain: osd-04, 并将该osd进程实例添加至osd-04中

```
[root@node6-4 ~]# ceph osd crush add-bucket osd-04 osd-domain

[root@node6-4 ~]# ceph osd crush add osd.9 0.024 osd-domain=osd-04

add item id 9 name 'osd.9' weight 0.024 at location {osd-domain=osd-04} to crush map

[root@node6-4 ~]# ceph osd crush add osd.10 0.024 osd-domain=osd-04

add item id 10 name 'osd.10' weight 0.024 at location {osd-domain=osd-04} to crush map

[root@node6-4 ~]# ceph osd crush add osd.11 0.024 osd-domain=osd-04

add item id 11 name 'osd.11' weight 0.024 at location {osd-domain=osd-04} to crush map
```

3.将osd-04移动至replica-02中

```
[root@node6-4 ~]# ceph osd crush move osd-04 replica-domain=replica-02
moved item id -16 name 'osd-04' to location {replica-domain=replica-02} in crush map
```

4.新建failure-domain: sata-02, 并将replica-02移动至sata-02中

```
[root@node6-4 ~]# ceph osd crush add-bucket sata-02 failure-domain

added bucket sata-02 type failure-domain to crush map

[root@node6-4 ~]# ceph osd crush move replica-02 failure-domain=sata-02

moved item id -15 name 'replica-02' to location {failure-domain=sata-02} in crush map
```

5.修改crush map文件

```
[root@node6-4 ~]# ceph osd getcrushmap -o /tmp/001_old_map.bin

got crush map from osdmap epoch 145

[root@node6-4 ~]# crushtool -d /tmp/001_old_map.bin -o /tmp/001_old_map.txt

[root@node6-4 ~]# vim /tmp/001_old_map.txt ## 修改crush map文件, 添加如下内容
>>

rule sata-02 {
ruleset 7
type replicated
min_size 1
max_size 10
step take sata-02
step choose firstn 1 type replica-domain
step chooseleaf firstn 0 type osd-domain
step emit
}
```

<<

```
[root@node6-4 ~]# crushtool -c /tmp/001_old_map.txt -o /tmp/001_new.map.bin
```

```
[root@node6-4 ~]# ceph osd setcrushmap -i /tmp/001_new.map.bin
set crush map
```

6.查看新的crush map

```
[root@node6-4 ~]# ceph osd tree
ID  WEIGHT  TYPE NAME                     UP/DOWN REWEIGHT PRIMARY-AFFINITY
-17 0.07196 failure-domain sata-02
-15 0.07196     replica-domain replica-02
-16 0.07196         osd-domain osd-04
  9 0.02399             osd.9                  up  1.00000          1.00000
 10 0.02399             osd.10                 up  1.00000          1.00000
 11 0.02399             osd.11                 up  1.00000          1.00000
-12 0.21599 failure-domain sata-01
 -8 0.21599     replica-domain replica-01
 -9 0.07199         osd-domain osd-01
  0 0.02399             osd.0                  up  1.00000          1.00000
  1 0.02399             osd.1                  up  1.00000          1.00000
  2 0.02399             osd.2                  up  1.00000          1.00000
-10 0.07199         osd-domain osd-02
  3 0.02399             osd.3                  up  1.00000          1.00000
  4 0.02399             osd.4                  up  1.00000          1.00000
  5 0.02399             osd.5                  up  1.00000          1.00000
-11 0.07199         osd-domain osd-03
  6 0.02399             osd.6                  up  1.00000          1.00000
  7 0.02399             osd.7                  up  1.00000          1.00000
  8 0.02399             osd.8                  up  1.00000          1.00000
 -1 0.29095 root default
 -5 0.07300     rack rack-01
 -2 0.07300         host node6-1
  0 0.02399             osd.0                  up  1.00000          1.00000
  1 0.02399             osd.1                  up  1.00000          1.00000
  2 0.02399             osd.2                  up  1.00000          1.00000
 -6 0.07300     rack rack-02
 -3 0.07300         host node6-2
  3 0.02399             osd.3                  up  1.00000          1.00000
  4 0.02399             osd.4                  up  1.00000          1.00000
  5 0.02399             osd.5                  up  1.00000          1.00000
 -7 0.07300     rack rack-03
 -4 0.07300         host node6-3
  6 0.02399             osd.6                  up  1.00000          1.00000
  7 0.02399             osd.7                  up  1.00000          1.00000
  8 0.02399             osd.8                  up  1.00000          1.00000
-14 0.07196     rack rack-04
-13 0.07196         host node6-4
  9 0.02399             osd.9                  up  1.00000          1.00000
 10 0.02399             osd.10                 up  1.00000          1.00000
 11 0.02399             osd.11                 up  1.00000          1.00000
```

至此向ceph集群中添加osd主机的过程已经操作完成