

ceph集群中新增osd节点和mon节点

ceph集群中新增osd主机节点

前提条件:

1. 需要在新加入的主机new7-1上安装ceph所需要的相关程序包

具体过程请参考 ==> [手工创建ceph集群过程中的"ceph集群部署前提"相关步骤](#)

2. `[root@new7-1 ~]# hostnamectl --static --transient set-hostname new7-1` `## 将主机名称统一修改成new7-1`

3. 修改ceph集群中的mon节点的主机的host列表文件添加new7-1的主机名称, 并将本地的公钥复制到新加入主机, 保证mon节点能够与new7-1进行无密码通信

部署过程:

1. 在ceph集群中新建osd id, 该id就作为新添加的osd的id

```
[root@node2 ~]# ceph osd create                ##
在ceph集群中的某个mon节点上新建一个新的osd-number, 该osd-number就会作为新加入的osd节点的编号
4
```

2. 对新建osd创建指定的磁盘分区

```
[root@new7-1 ~]# mkfs.xfs -f /dev/vdb1
## 在new7-1主机上新建分区用于作为本机osd的分区, 对该分区进行格式化
meta-data=/dev/vdb1                isize=256     agcount=4, agsize=65536 blks
             =                        sectsz=512     attr=2, projid32bit=1
             =                        crc=0          finobt=0
data        =                        bsize=4096    blocks=262144, imaxpct=25
             =                        sunit=0        swidth=0 blks
naming      =version 2                bsize=4096    ascii-ci=0 ftype=0
log         =internal log            bsize=4096    blocks=2560, version=2
             =                        sectsz=512    sunit=0 blks, lazy-count=1
realtime    =none                     extsz=4096    blocks=0, rtextents=0
[root@new7-1 ~]# mkdir -pv /var/lib/ceph/osd/ceph-4
## 新建osd的默认工作目录, 其中目录名称中的4就是通过mon节点产生的osd-number
[root@new7-1 ~]# mount /dev/vdb1 /var/lib/ceph/osd/ceph-4
## 将默认工作目录挂载至新建的分区上
```

3. 在new7-1节点上添加mon指定的key

```
[root@new7-1 ~]# vim /etc/ceph/ceph.client.admin.keyring
>>
[client.admin]
             key = AQBri6lWdn+FCBAAsizVnLm3n4HDPXja657UgA==
## 将key修改为mon节点上对应配置文件的key
<<
```

4. 在new7-1节点初始化osd的工作目录

先将mon节点中的ceph配置目录下的配置文件ceph.conf复制到new7-1节点

```
[root@new7-1 ~]# ceph-osd -i 4 --mkfs --mkkey ## 第二次执行该命令时，能够正确执行，并产生全部生成文件
2016-02-01 18:03:55.284890 7f0516ac4880 -1 journal FileJournal::_open: disabling aio for non-block
journal. Use journal_force_aio to force use of aio anyway
2016-02-01 18:03:55.284938 7f0516ac4880 -1 journal check: ondisk fsid
00000000-0000-0000-0000-000000000000 doesn't match expected 4fe69612-55ea-442d-bc8c-d217670d8c12,
invalid (someone else's?) journal
2016-02-01 18:03:55.325926 7f0516ac4880 -1 journal FileJournal::_open: disabling aio for non-block
journal. Use journal_force_aio to force use of aio anyway
2016-02-01 18:03:55.329910 7f0516ac4880 -1 filestore(/var/lib/ceph/osd/ceph-4) could not find
23c2fcde/osd_superblock/0//1 in index: (2) No such file or directory
2016-02-01 18:03:55.414299 7f0516ac4880 -1 created object store /var/lib/ceph/osd/ceph-4 journal
/var/lib/ceph/osd/ceph-4/journal for osd.4 fsid e9a656ca-29ec-4d07-b65c-4e0e82674f36
2016-02-01 18:03:55.414355 7f0516ac4880 -1 auth: error reading file: /var/lib/ceph/osd/ceph-4/keyring:
can't open /var/lib/ceph/osd/ceph-4/keyring: (2) No such file or directory
2016-02-01 18:03:55.414461 7f0516ac4880 -1 created new key in keyring /var/lib/ceph/osd/ceph-4/keyring
[root@new7-1 ~]# ll /var/lib/ceph/osd/ceph-4/
total 40
-rw-r--r-- 1 root root      37 Feb  1 18:03 ceph_fsid
drwxr-xr-x 4 root root      61 Feb  1 18:03 current
-rw-r--r-- 1 root root      37 Feb  1 18:02 fsid
-rw-r--r-- 1 root root 1073741824 Feb  1 18:03 journal
-rw----- 1 root root      56 Feb  1 18:03 keyring
-rw-r--r-- 1 root root      21 Feb  1 18:03 magic
-rw-r--r-- 1 root root       6 Feb  1 18:03 ready
-rw-r--r-- 1 root root       4 Feb  1 18:02 store_version
-rw-r--r-- 1 root root      53 Feb  1 18:02 superblock
-rw-r--r-- 1 root root       2 Feb  1 18:03 whoami
```

5. 注册新的osd认证密钥

```
[root@new7-1 ~]# ceph auth add osd.4 osd 'allow *' mon 'allow rwx' -i /var/lib/ceph/osd/ceph-4/keyring
added key for osd.4
```

6. 将new7-1节点添加至crush map中

```
[root@new7-1 ~]# ceph osd crush add-bucket osd4 host
added bucket osd4 type host to crush map
[root@new7-1 ~]# ceph osd crush move osd4 root=default
moved item id -8 name 'osd4' to location {root=default} in crush map
[root@new7-1 ~]# ceph osd crush create-or-move osd.4 2.0 root=default host=osd4
create-or-move updating item name 'osd.4' weight 2 at location {host=osd4,root=default} to crush map
```

7. 在mon节点上修改ceph的配置文件/etc/ceph/ceph.conf，添加对于新添加的osd.4节点的配置

```
[root@node2 ~]# vim /etc/ceph/ceph.conf ## 需要将该配置文件复制到ceph集群中的所有主机节点中去
>>
[osd.4]
host = new7-1
devs = /dev/vdb1
<<
```

8. 在new7-1节点上启动osd进程

```
[root@new7-1 ~]# /etc/init.d/ceph start osd.4
=== osd.4 ===
Mounting xfs on new7-1:/var/lib/ceph/osd/ceph-4
create-or-move updating item name 'osd.4' weight 0 at location {host=new7-1,root=default} to crush map
Starting Ceph osd.4 on new7-1...
Running as unit run-2771.service.
[root@new7-1 ~]# ceph -s
  cluster e9a656ca-29ec-4d07-b65c-4e0e82674f36
    health HEALTH_WARN
      9 pgs stale
      9 pgs stuck stale
      1 mons down, quorum 0,2 node2,node3
    monmap e2: 3 mons at
{node1=192.168.5.31:6789/0,node2=192.168.5.30:6789/0,node3=192.168.5.32:6789/0}
      election epoch 8, quorum 0,2 node2,node3
    osdmap e52: 6 osds: 4 up, 4 in
    pgmap v971: 64 pgs, 1 pools, 25297 kB data, 33 objects
      3302 MB used, 9969 MB / 13272 MB avail
        55 active+clean
        9 stale+active+clean
[root@new7-1 ~]# ceph osd tree
ID WEIGHT  TYPE NAME                UP/DOWN REWEIGHT PRIMARY-AFFINITY
-6 1.00000  root default
-5      0    host osd5
-7      0    host node5
-8      0    host osd4
-9 1.00000  host new7-1
  4 1.00000      osd.4          up  1.00000          1.00000
-4 3.00000  host node3
  2 3.00000      osd.2          up  1.00000          1.00000
-3 4.00000  host node2
  1 2.00000      osd.1          up  1.00000          1.00000
  5 2.00000      osd.5          up  1.00000          1.00000
-2 1.00000  host node1
  0 1.00000      osd.0         down           0          1.00000
-1      0  host new7-2
  3      0  osd.3          down           0          1.00000
```

ceph集群中删除osd主机节点

部署过程(需要在mon节点主机上进行)

1. 在ceph中的crush算法找哦你删除指定osd节点

```
[root@node2 ~]# ceph osd tree
ID WEIGHT  TYPE NAME          UP/DOWN REWEIGHT PRIMARY-AFFINITY
... ..
 0 1.00000  osd.0             down      0          1.00000
-1 1.50000 host new7-2
 3 1.50000  osd.3             up    1.00000          1.00000
[root@node2 ~]# ceph osd crush remove osd.0
removed item id 0 name 'osd.0' from crush map
[root@node2 ~]# ceph osd tree
ID WEIGHT  TYPE NAME          UP/DOWN REWEIGHT PRIMARY-AFFINITY
... ..
 0          0 osd.0             down      0          1.00000
```

2. 在ceph集群中删除指定osd节点的认证关系

```
[root@node2 ~]# ceph auth del osd.0
2016-02-02 11:11:33.402822 7f407bd77700 0 -- :/1029728 >> 192.168.5.31:6789/0 pipe(0x7f4080065010 sd=3
:0 s=1 pgs=0 cs=0 l=1 c=0x7f408005ba10). fault
updated
```

3. 从ceph集群中删除指定osd节点

```
[root@node2 ~]# ceph osd rm osd.0
removed osd.0
[root@node2 ~]# ceph osd tree
ID WEIGHT TYPE NAME UP/DOWN REWEIGHT PRIMARY-AFFINITY
此时已经删掉了osd.0的相关信息
```

4. 修改ceph的配置文件/etc/ceph/ceph.conf，删除其中的关于osd.0节点的配置字段，并将该配置文件同步至ceph集群中的全部节点

```
[root@node2 ~]# vim /etc/ceph/ceph.conf ## 删除指定osd的相关字段
>>
[osd.1]
  host = x.x.x.x
<<
```

ceph集群中新增mon节点主机

前提条件:

1. 需要在新加入的主机new7-1上安装ceph所需要的相关程序包

具体过程请参考 ==>[手工创建ceph集群过程中的"ceph集群部署前提"相关步骤](#)

2. `[root@new7-1 ~]# hostnamectl --static --transient set-hostname new7-1` ## 将主机名称统一修改成new7-1

3. 修改ceph集群中的mon节点的主机的host列表文件添加new7-1的主机名称, 并将本地的公钥复制到新加入主机, 保证mon节点能够与new7-1进行无密码通信

部署过程

1. 创建mon节点需要的目录

```
[root@new-node7-4 ~]# mkdir /var/lib/ceph/mon/ceph-new-node7-4
```

2. 从ceph集群中原有的mon节点获取keyring文件

```
[root@new-node7-4 ~]# scp node7-1:/tmp/ceph.mon.keyring /tmp/ ## 将mon节点的keyring文件复制到本地
[root@new-node7-4 ~]# ceph auth get mon. -o /tmp/ceph.mon.keyring
exported keyring for mon.
```

3. 获取mon节点的map信息

```
[root@new-node7-4 ~]# ceph mon getmap -o /tmp/ceph.mon.map
got monmap epoch 1
```

4. 创建mon的数据目录

```
[root@new-node7-4 ~]# ceph-mon -i new-node7-4 --mkfs --monmap /tmp/ceph.mon.map --keyring
/tmp/ceph.mon.keyring
ceph-mon: set fsid to 2cfff3d5-77fc-4ce4-88be-3bfc4c2b6d84
ceph-mon: created monfs at /var/lib/ceph/mon/ceph-new-node7-4 for mon.new-node7-4
```

5. 将新的mon节点添加至ceph集群的mon列表

```
[root@new-node7-4 ~]# ceph mon add new-node7-4 192.168.5.46:6789
added mon.new-node7-4 at 192.168.5.46:6789/0
```

6. 启动新添加的mon进程

```
[root@new-node7-4 ~]# scp node7-1:/var/lib/ceph/mon/ceph-node7-1/keyring
/var/lib/ceph/mon/ceph-new-node7-4/ ## 复制ceph集群中的mon节点的keyring文件到本地
[root@new-node7-4 ~]# ceph-mon -i new-node7-4 --public-addr 192.168.5.46:6789 ##
通过--public-addr选项用于指定mon守护进程绑定的地址
```

7. 验证新添加的mon的运行状态

```
[root@new-node7-4 ~]# ceph mon stat
e2: 3 mons at {new-node7-4=192.168.5.46:6789/0,node7-1=192.168.5.43:6789/0,node7-2=192.168.5.45:6789/0},
election epoch 16, quorum 0,1,2 node7-1,node7-2,new-node7-4
```