ceph-deploy工具创建ceph集群

系统版本
系统版本：centos6.6
内核版本：2.6.32-504.el6.x86_64

ceph集群拓扑说明

IP地址	节点名称	功能角色
10.10.10.161	node161	deploy
10.10.10.162	node162	mon&mds
10.10.10.163	node163	mds
10.10.10.164	node164	mds

---

ceph集群部署前提

1. 关闭各个node的iptables和SElinux
1-1. 关闭SElinux
```
# setenforce 0    ## 通过指定关闭SELinux
# vim /etc/selinux/config    ## 修改配置文件/etc/selinux/config中的SELINUX字段
>>
SELINUX=disabled
<<
```

1-2. 关闭iptables
```
# chkconfig iptables off
# chkconfig --off iptables
```

如果iptables或SELinux仍然与运行，建议将主机重启

---

2. 添加hosts文件实现集群内主机与主机之间可以通过名称相互能够解析
在各个node主机上编辑host配置文件，添加如下内容
```
# vim /etc/hosts
>>
10.10.10.161 node161
10.10.10.162 node162
10.10.10.163 node163
10.10.10.164 node164
<<
```
此处应该保证hostname应该与此处的设置的节点名称应该相同

---
3. 在node161上添加yum的base源和epel源(此处添加的是阿里云的yum源)
```
[root@node161 ~]# mv /etc/yum.repos.d{,.bak}
[root@node161 ~]# mkdir /etc/yum.repos.d
```

3-1. `[root@node161 ~]# wget -O /etc/yum.repos.d/CentOS-Base.repo http://mirrors.aliyun.com/repo/Centos-6.repo` ## 获取yum的base源的配置文件

3-2. `[root@node161 ~]# wget -O /etc/yum.repos.d/epel.repo http://mirrors.aliyun.com/repo/epel-6.repo` ## 获取yum的epel源配置文件

```
[root@ceph-node-1 ~]# ll /etc/yum.repos.d/
total 8
-rw-r--r-- 1 root root 2572 May 15 2015 CentOS-Base.repo
-rw-r--r-- 1 root root 1083 May 15 2015 epel.repo
```
---
4. 将作为ceph-deploy的node161的公钥复制到其他各个节点，实现node161无密码登录其他各个节点

```
[root@node161 ~]# yum install -y openssh-server    ## 安装openssh-server

[root@node161 ~]# ssh-keygen -t rsa    ## 在本地生成ssh密钥
```

[root@node161 ~]# for no in {161..164}; do ssh-copy-id -i /root/.ssh/id_rsa.pub root@node${no};done   ##
将node161的公钥复制到其他各个节点

[root@node161 ~]# for no in {161..164}; do ssh node${no} hostname; done   ##
使用for循环显示远程主机名称验证无密码登录是否配置成功

---
5. node161添加ceph的yum源的repo配置文件
[root@node161 ~]# vim /etc/yum.repos.d/ceph.repo
>>
[ceph]
name=Ceph packages for $basearch
baseurl=http://ceph.com/rpm/rhel6/$basearch
enabled=1
gpgcheck=0
type=rpm-md
gpgkey=https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc

[ceph-noarch]
name=Ceph noarch packages
baseurl=http://ceph.com/rpm/rhel6/noarch
enabled=1
gpgcheck=0
type=rpm-md
gpgkey=https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc

[ceph-source]
name=Ceph source packages
baseurl=http://ceph.com/rpm/rhel6/SRPMS
enabled=0
gpgcheck=0
type=rpm-md
gpgkey=https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc
<<

[root@node161 ~]# for no in {162..164};do ssh node${no} mv /etc/yum.repos.d{,.bak};ssh node${no} mkdir /etc/yum.repos.d;done
[root@node161 ~]# for no in {162..164};do scp /etc/yum.repos.d/* node${no}:/etc/yum.repos.d/;done
---
6. 配置ntp服务，开启时间服务，保证集群服务器时间统一；

---
7. 安装ceph集群所依赖的程序包
[root@node161 ~]# for no in {161..164};do echo "===node${no}===";ssh node${no} yum -y install *argparse* redhat-lsb xfs* rsync;done
因为ceph集群的依赖程序包较为复杂，建议使用该种方法来安装ceph所依赖的程序包，以免在后续过程中出现错误

8. 在ceph集群中的各个节点创建指定的ceph数据目录/data/ceph
[root@node161 ~]# for no in {161..164}; do echo "===node${no}==="; ssh node${no} mkdir -pv /data/ceph;done

9. 在node162~node164主机上创建新的分区作为ceph的存储分区

---

deploy节点部署

1.在node161节点上安装ceph-deploy工具

[root@node161 ~]# cd /data/ceph

[root@node161 ceph]# yum update -y

[root@node161 ceph]# yum install ceph-deploy -y   ## 在node161上安装ceph-deploy

2. 如果之前在各个节点上部署过ceph集群时，需要执行一下操作来清除之前的数据

[root@node161 ceph]# stop ceph-all   ## 在node161上停止所有ceph进程

[root@node161 ceph]# ceph-deploy uninstall node161 node162 node163 node164   ## 在所有node上卸载所有ceph程序包

```
[root@node161 ceph]# ceph-deploy purge node161 node162 node163 node164     ## 在所有node上清除所有数据
[root@node161 ceph]# ceph-deploy purgedata node161 node162 node163 node164     ## 在所有node上清除所有元数据
[root@node161 ceph]# ceph-deploy forgetkeys     ## 在node161上清除所有key
```

3.清除所有节点的无用文件，并创建配置文件目录

```
[root@node161 ceph]# for no in {161..164};do echo "===node${no}===";ssh node${no} rm -rf /etc/ceph/*;done
[root@node161 ceph]# for no in {161..164};do echo "===${no}===";ssh node${no} mkdir -pv /etc/ceph/;done
```

---

monitor节点部署

1. monitro节点部署

```
[root@node161 ceph]# ceph-deploy new node162     ## 新建一个ceph集群，并将node162单个节点作为该集群的monitor节点
[root@node161 ceph]# ll     ## 查看该数据目录下新产生的文件
-rw-r--r-- 1 root root 229 Jan 19 00:40 ceph.conf
-rw-r--r-- 1 root root 3791 Jan 19 00:40 ceph.log
-rw------- 1 root root 73 Jan 19 00:40 ceph.mon.keyring
```

2.在deploy节点上向所有node安装ceph软件

```
[root@node161 ceph]# ceph-deploy install node16{1..4}
## 该命令会持续较长时间
```

3.添加初始monitor节点和收集密钥

```
[root@node161 ceph]# ceph-deploy --overwrite-conf mon create-initial   ## 此时node162就成为了monitor节点
[root@node161 ceph]# ls   ## 查看此时新增的文件
ceph.bootstrap-mds.keyring ceph.bootstrap-osd.keyring ceph.bootstrap-rgw.keyring
[root@node161 ceph]# ceph-deploy gatherkeys node162
```

4.node161向其他node同步本地的ceph配置文件

```
[root@node161 ceph]# for no in {161..164};do echo "===node${no}===";rsync -avp --delete /data/ceph node${no}:/data/;done
```

5.查看node162上的monitor进程是否正常运行

```
[root@node162 ceph]# ps -ef | grep ceph
root 3354 1 0 02:31 ? 00:00:00 /usr/bin/ceph-mon -i node162 --pid-file /var/run/ceph/mon.node162.pid -c /etc/ceph/ceph.conf
--cluster ceph
root 3839 3131 6 02:55 pts/0 00:00:00 grep --color=auto ceph
```

OSD节点部署

1.在node161上为各个osd节点添加设备

```
## 在ceph集群中需要使用投票机制，因此ceph集群中的osd节点数量应该为单数
[root@node161 ceph]# ceph-deploy --overwrite-conf osd prepare node162:/dev/sda5
[root@node161 ceph]# ceph-deploy --overwrite-conf osd prepare node163:/dev/sda5
[root@node161 ceph]# ceph-deploy --overwrite-conf osd prepare node164:/dev/sda5
```

2. 在node161上为各个osd节点激活设备

```
[root@node161 ceph]# ceph-deploy osd activate node162:/dev/sda5

[root@node161 ceph]# ceph-deploy osd activate node163:/dev/sda5

[root@node161 ceph]# ceph-deploy osd activate node164:/dev/sda5
```

3. 复制node161的ceph配置文件及密钥到mon和osd节点

```
[root@node161 ceph]# ceph-deploy admin node161 node162 node163 node164
```

4. 对node161上的/etc/ceph/ceph.client.admin.keyring文件添加其他用户的读权限

```
[root@node161 ceph]# chmod +r /etc/ceph/ceph.client.admin.keyring

[root@node161 ceph]# ll /etc/ceph/ceph.client.admin.keyring

-rwx--x--x 1 root root 63 Jan 19 06:09 /etc/ceph/ceph.client.admin.keyring
```

5. 查看三台监控节点的选举状态和集群的健康状态

```
[root@node162 ceph]# ceph quorum_status --format json-pretty

{
    "election_epoch": 2,
    "quorum": [
        0
    ],
    "quorum_names": [
        "node162"
    ],
    "quorum_leader_name": "node162",
    "monmap": {
        "epoch": 1,
        "fsid": "d4599825-901e-47a9-8bf8-4cb32f797832",
        "modified": "0.000000",
        "created": "0.000000",
        "mons": [
            {
                "rank": 0,
                "name": "node162",
                "addr": "10.10.10.162:6789\/0"
            }
        ]
    }
}


[root@node162 ceph]# ceph health

HEALTH_OK
```

在CentOS7上用ceph-deploy工具创建ceph集群

在Ustack公有云搭建ceph测试环境。

从零开始源码搭建ceph集群给出了一个很好的从零开始搭建ceph集群的方法，但是公有云上机器编译ceph比较慢，所以提供另外一种使用二进制安装包搭建ceph集群的方法。如需使用自己编译的ceph安装，可以打包成rpm包，建立yum repo替换ceph官方源进行安装。

本文档以CentOS7.0为例搭建ceph集群

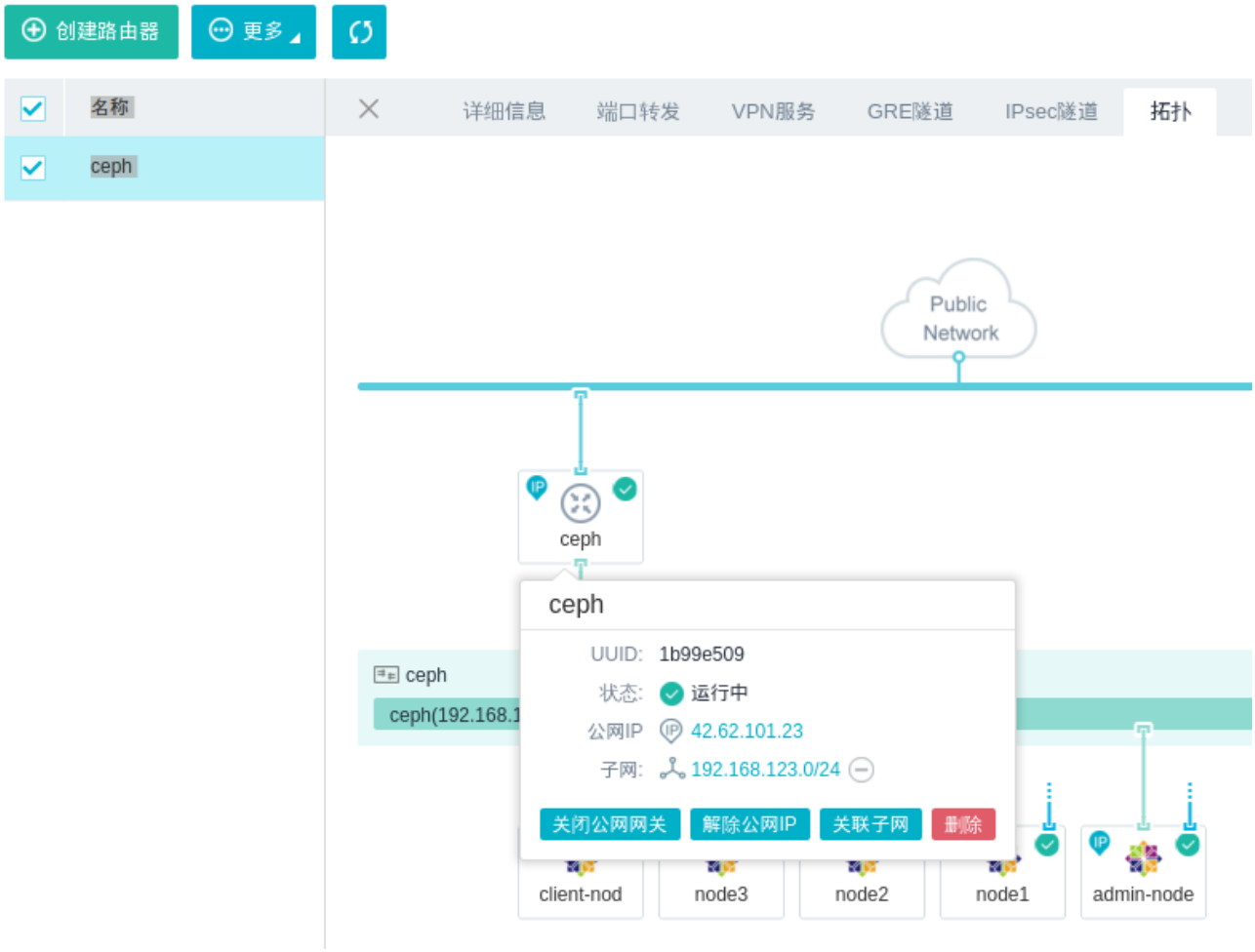
0. 目的：在Ustack公有云平台上使用CentOS 7.0搭建ceph集群环境。

   网络规划：10.250.10.*/22 管理网络， 192.168.123.1/24数据网络

1. 环境准备

1.1 网络

在公有云上创建CentOS7.0虚拟机。Ustack公有云提供的公网地址有限制，所以需要创建路由和网络并将二者进行关联。在路由器上绑定公网IP，所以子网内的所有可以访问公网并使用yum安装搭建环境所需的软件包。新建5个虚拟网卡，选择自动分配地址，以备在主机准备环节加载到虚拟机，用于虚拟机与公网通信。



1.2 主机

新建虚拟机admin-node并添加外网地址，进行以下配置：

1.2.1 增加 yum配置文件

```
#vim /etc/yum.repos.d/ceph-deploy.repo
```

添加以下内容：

```
[ceph-noarch]
name=Ceph noarch packages
baseurl=http://ceph.com/rpm-firefly/el7/noarch
enabled=1
gpgcheck=1
type=rpm-md
gpgkey=https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc
```

1.2.2 安装软件：

```
#yum update && yum installntp ntpupdate ntp-docyum-plugin-priorities
```

1.2.3 关闭防火墙和SELinux

```
#systemctl stop firewall.service

#sudo setenforce 0
```

1.2.4 创建ceph用户并设置密码

```
#sudo adduser -d /home/ceph -m ceph
# passwd ceph

#echo "ceph ALL = （root） NOPASSWD:ALL" >> /etc/sudoers.d/ceph
```

chmod 0440 /etc/sudoers.d/ceph

执行命令visudo修改suoders文件：

把Defaults requiretty 这一行修改为修改 Defaults:ceph ! requiretty

如果不进行修改ceph-depoy利用ssh执行命令将会出错

1.2.5 创建主机快照，并使用主机快照新建名为node1, node2, node3,
client-node的虚拟机。ceph-deploy工具都是通过主机名与其他节点通信。在admin-node上使用root登录每台虚拟机并修改主机名，修改主机名的命令为：

hostnamectl set-hostname ${NEWNAME}

在admin-node上添加ip地址和主机名对应关系到/etc/hosts。

192.168.123.104 node1

192.168.123.105node2

192.168.123.106node3

192.168.123.107client-node

1.2.6 外网通信配置

所有虚拟主机admin-node, node1, node2, node3, client-node分别加载section 1.1中建立的虚拟网卡。

需要注意的是node1, node2, node3, client-node添加虚拟网卡后，主机会添加一条默认路由到系统路由表，需要删掉系统原有的ustack内网默认路由，这样使访问外网的流量通过Section1.1中新建的路由连接外网。admin-node上的路由表可以不用改变。

ip r s

default via 10.250.8.1 dev eth0 proto static metric 100 ←-删掉这条

default via 192.168.123.1 dev eth1 proto static metric 101

10.250.8.0/22 dev eth0 proto kernel scope link src 10.250.10.243 metric 100

192.168.123.0/24 dev eth1 proto kernel scope link src 192.168.123.103 metric 100

ip r del default via 10.250.8.1 dev eth0 proto static metric 100

ip r s

default via 192.168.123.1 dev eth1 proto static metric 101

10.250.8.0/22 dev eth0 proto kernel scope link src 10.250.10.243 metric 100

192.168.123.0/24 dev eth1 proto kernel scope link src 192.168.123.103 metric 100

1.2.7 配置admin-node使用ssh自动登录node1, node2, node3, client-node。
在admin-node 上：

su - ceph

以下使用ceph账户执行：
$ ssh-keygen

$ ssh-copy-id ceph@node1

$ ssh-copy-id ceph@node2

$ ssh-copy-id ceph@node3

$ ssh-copy-id ceph@client-node

1.2.8 在admin-root上为ceph用户添加./ssh/config文件，并修改文件权限为0600：

$ cat ~/.ssh/config

Host node1

Hostname 192.168.1.104

User ceph

Host node2

Hostname 192.168.1.105

User ceph

Host node3

Hostname 192.168.1.106

User ceph

Host client-node

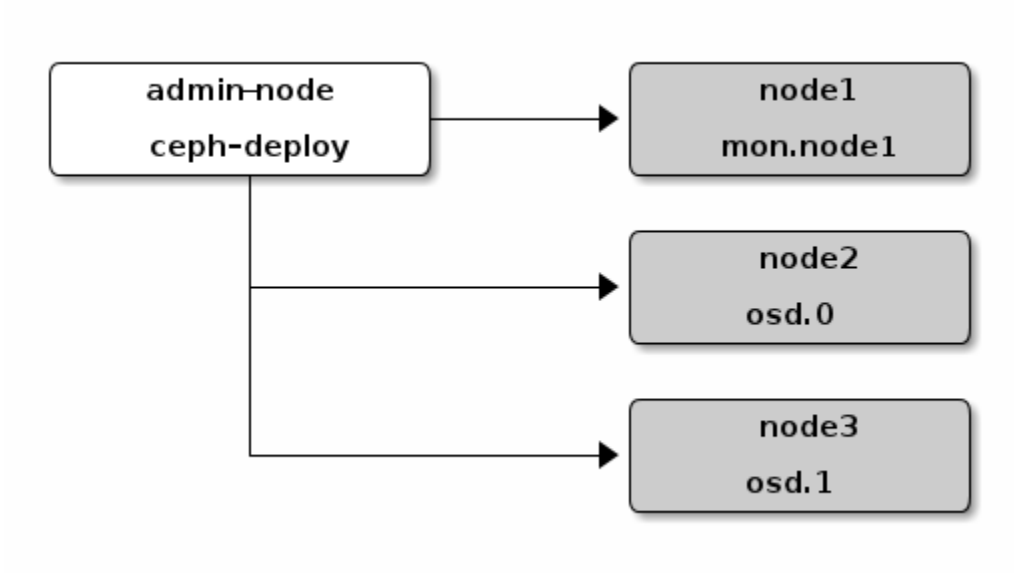
Hostname 192.168.1.107

User ceph

$ chmod 0600 ~/.ssh/config

2. ceph部署

下面我在准备好的环境中进行ceph部署：



2.1 在admin-node使用yum安装ceph-deploy

$ sudo yum installceph-deploy

2.2执行以下命令创建以node1为监控节点的集群。

$ ceph-deploy new node1

执行该命令后将在当前目录生产ceph.conf文件，打开文件并增加一下内容：

$ osd pool default size = 2

2.3利用ceph-deploy为节点安装ceph

$ ceph-deploy install admin-node node1 node2 node3

2.4初始化监控节点并收集keyring：

$ ceph-deploy mon create-initial

为存储节点osd进程分配磁盘空间：

$ ssh node2mkdir /var/local/osd0

$ ssh node3mkdir /var/local/osd1

2.5接下来通过admin-node节点的ceph-deploy开启其他节点osd进程，并激活。

$ ceph-deploy osd prepare node2:/var/local/osd0 node3:/var/local/osd1

$ ceph-deploy osd activate node2:/var/local/osd0 node3:/var/local/osd1

把admin-node节点的配置文件与keyring同步至其它节点：

$ ceph-deploy admin admin-node node1 node2 node3

```
$ sudo chmod +r /etc/ceph/ceph.client.admin.keyring
```

2.6最后通过命令查看集群健康状态：

```
$ ceph health
```

如果成功将提示：HEALTH_OK

```
$ ceph quorum_status --format json-pretty
{
"election_epoch": 2,
"quorum": [
0
],
"quorum_names": [
"node1"
],
"quorum_leader_name": "node1",
"monmap": {
"epoch": 1,
"fsid": "fc1aa390-ebcc-4a33-8909-b150ed2638c4",
"modified": "0.000000",
"created": "0.000000",
"mons": [
{
"rank": 0,
"name": "node1",
"addr": "10.250.10.25:6789\/0"
}
]
}
}
```