### 第五十章：Ceph分布式存储

**一、Ceph概述；**

**二、基本架构；**

**三、架构组件详解；**

**四、Ceph数据存储过程；**

**五、Ceph的优势；**

**六、案例：部署Ceph分布式存储集群；**

**一、Ceph概述：**

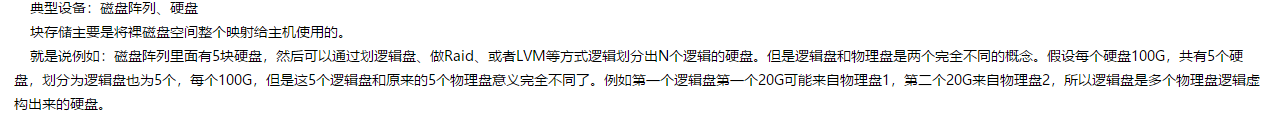
**概述：**Ceph是根据加州大学Santa Cruz分校的Sage Weil的博士论文所设计开发的新一代自由软件分布式文件系统，其设计目标是良好的可扩展性（PB级别以上）、高性能、高可靠性。Ceph其命名和UCSC（Ceph 的诞生地）的吉祥物有关，这个吉祥物是“Sammy”，一个香蕉色的蛞蝓，就是头足类中无壳的软体动物。这些有多触角的头足类动物，是对一个分布式文件系统高度并行的形象比喻。 其设计遵循了三个原则：数据与元数据的分离，动态的分布式的元数据管理，可靠统一的分布式对象存储机制。

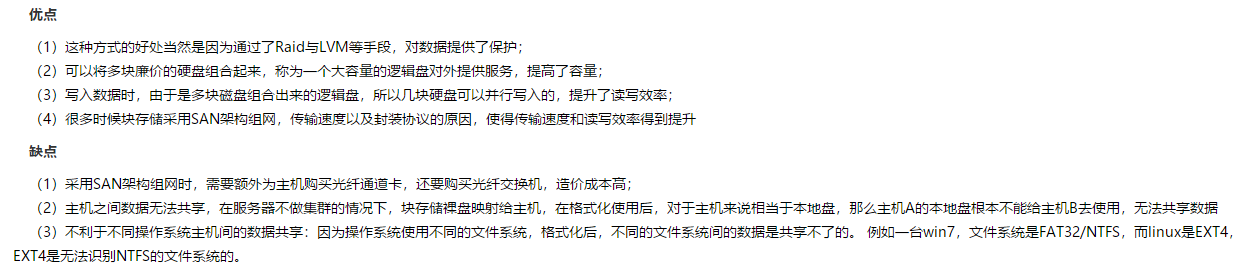
**二、基本架构：**

**1.Ceph**是一个高可用、易于管理、开源的分布式存储系统，可以在一套系统中同时提供对象存储、块存储以及文件存储服务。其主要由Ceph存储系统的核心RADOS以及块存储接口、对象存储接口和文件系统存储接口组成；

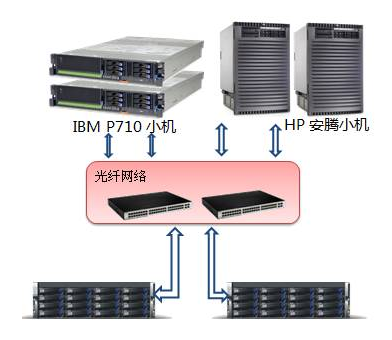
**2.存储类型：**

* **块存储：**

****

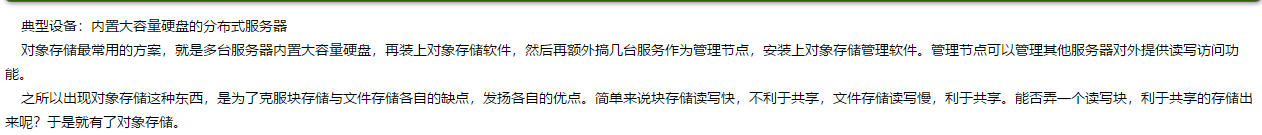
****

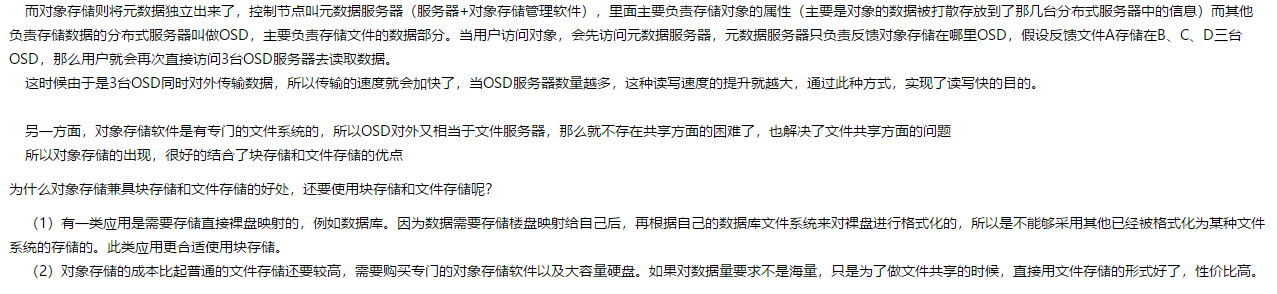
在常见的存储中 DAS、SAN 提供的也是块存储、openstack的cinder存储，例如iscsi的存储；



* **对象存储：**

对象存储概念出现得晚，存储标准化组织SINA早在2004年就给出了定义，但早期多出现在超大规模系统，所以并不为大众所熟知，相关产品一直也不温不火。一直到云计算和大数据的概念全民强推，才慢慢进入公众视野。前面说到的块存储和文件存储，基本上都还是在专有的局域网络内部使用，而对象存储的优势场景却是互联网或者公网，主要解决海量数据，海量并发访问的需求。基于互联网的应用才是对象存储的主要适配（当然这个条件同样适用于云计算，基于互联网的应用最容易迁移到云上，因为没出现云这个名词之前，他们已经在上面了），基本所有成熟的公有云都提供了对象存储产品，不管是国内还是国外；

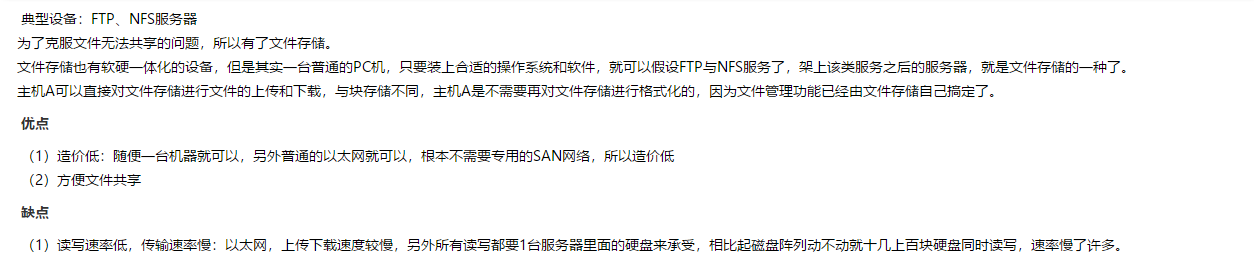




这种接口通常以 QEMU Driver 或者 Kernel Module 的方式存在，这种接口需要实现 Linux 的 Block Device 的接口或者 QEMU 提供的 Block Driver 接口，如 Swift 、S3 以及 Gluster、Sheepdog，AWS 的 EBS，青云的云硬盘和阿里云的盘古系统，还有 Ceph 的 RBD（RBD是Ceph面向块存储的接口）；



* **文件系统存储：**

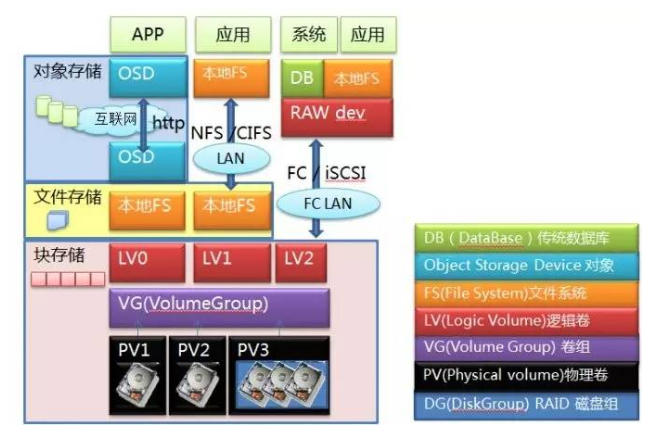
****

与传统的文件系统如 Ext4 是一个类型的，但区别在于分布式存储提供了并行化的能力，如 Ceph 的 CephFS (CephFS是Ceph面向文件存储的接口)，但是有时候又会把 GlusterFS ，HDFS 这种非POSIX接口的类文件存储接口归入此类。当然 NFS、NAS也是属于文件系统存储；

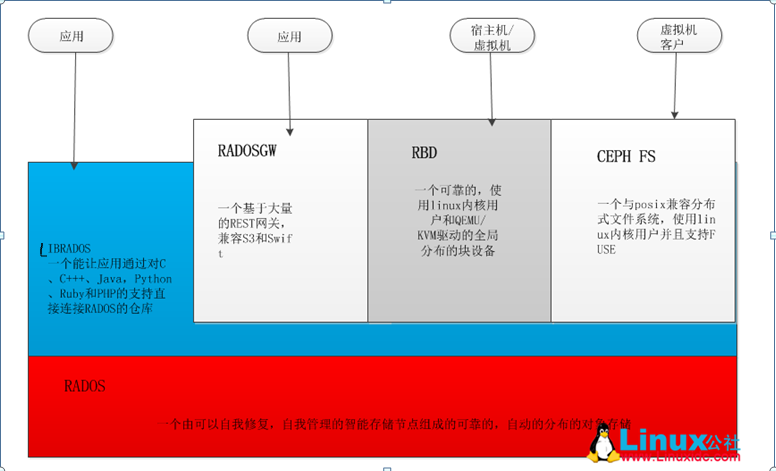
* **总结：对比；**

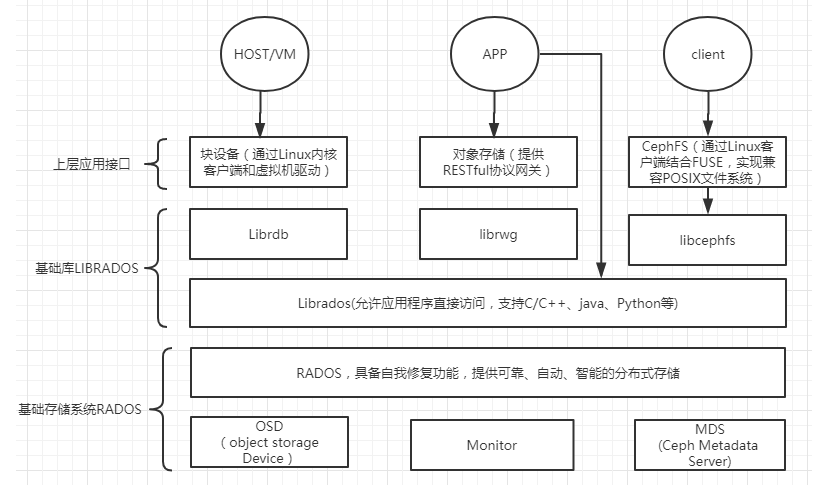


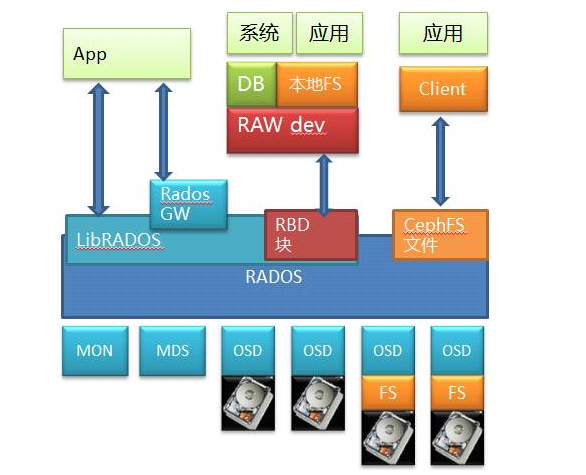




**3.Ceph基本架构：**



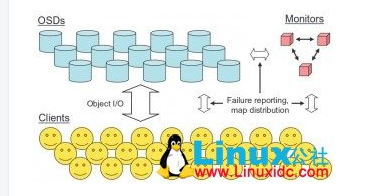


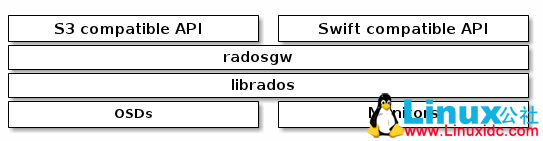
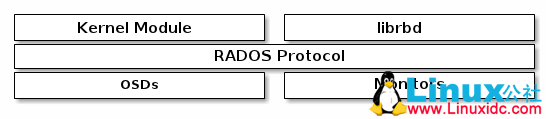


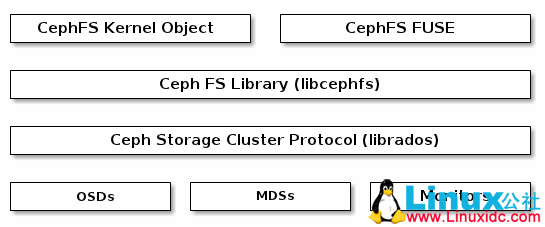
**三、架构组件详解：**

* **RADOS：**所有其他客户端接口使用和部署的基础。由以下组件组成：   
   OSD：Object StorageDevice，提供数据实体存储资源；  
   Monitor：维护整个Ceph集群中各个节点的心跳信息，维持整个集群的全局状态；  
   MDS：Ceph Metadata Server，文件系统元数据服务节点。MDS也支持多台机器分布 式的部署，以实现系统的高可用性。

典型的RADOS部署架构由少量的Monitor监控器以及大量的OSD存储设备组成，它能够在动态变化的基于异质结构的存储设备集群之上提供一种稳定的、可扩展的、高性能的单一逻辑对象存储接口。

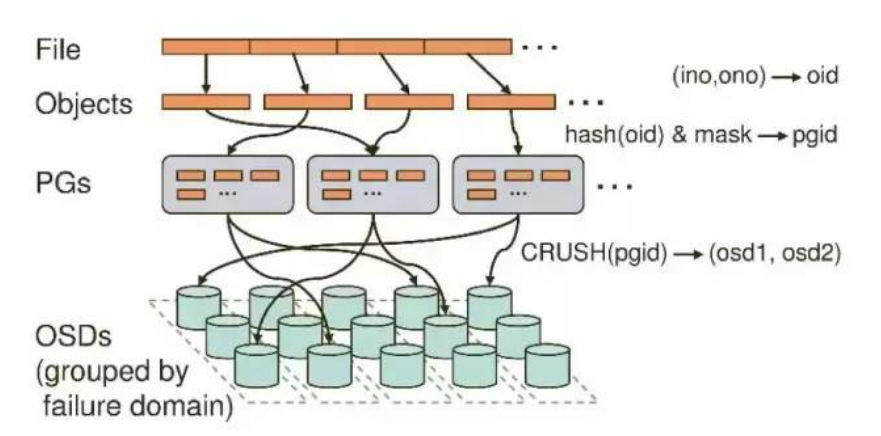


* **Ceph客户端接口(Clients) ：**Ceph架构中除了底层基础RADOS之上的LIBRADOS、RADOSGW、RBD以及Ceph FS统一称为Ceph客户端接口。简而言之就是RADOSGW、RBD以及Ceph FS根据LIBRADOS提供的多编程语言接口开发。所以他们之间是一个阶梯级过渡的关系。   
  **1.RADOSGW ：** Ceph对象存储网关，是一个底层基于librados向客户端提供RESTful接口的对象存储接口。目前Ceph支持两种API接口：   
  S3.compatible：S3兼容的接口，提供与Amazon S3大部分RESTfuI API接口兼容的API接口。   
  Swift.compatible：提供与OpenStack Swift大部分接口兼容的API接口。Ceph的对象存储使用网关守护进程（radosgw）, radosgw结构图如图所示：   
  **2.RBD ：**一个数据块是一个字节序列（例如，一个512字节的数据块）。基于数据块存储接口最常见的介质，如硬盘，光盘，软盘，甚至是传统的9磁道的磁带的方式来存储数据。块设备接口的普及使得虚拟块设备成为构建像Ceph海量数据存储系统理想选择。 在一个Ceph的集群中， Ceph的块设备支持自动精简配置，调整大小和存储数据。Ceph的块设备可以充分利用 RADOS功能，实现如快照，复制和数据一致性。Ceph的RADOS块设备（即RBD）通过RADOS协议与内核模块或librbd的库进行交互。。RBD的结构如图所示：   
  

**3.Ceph FS ：**Ceph文件系统（CEPH FS）是一个POSIX兼容的文件系统，使用Ceph的存 储集群来存储其数据。CEPH FS的结构图如下所示：   


**扩展理解地址：**https://www.sohu.com/a/144775333\_151779

**四、Ceph数据存储过程：**

Ceph存储集群从客户端接收文件，每个文件都会被客户端切分成一个或多个对象，然后将这些对象进行分组，再根据一定的策略存储到集群的OSD节点中，其存储过程如图所示：   
图中，对象的分发需要经过两个阶段的计算：  
1.对象到PG的映射。PG(PlaccmentGroup)是对象的逻辑集合。PG是系统向OSD节点分发数据的基本单位，相同PG里的对象将被分发到相同的OSD节点中(一个主OSD节点多个备份OSD节点)。对象的PG是由对象ID号通过Hash算法，结合其他一些修正参数得到的。   
2.PG到相应的OSD的映射，RADOS系统利用相应的哈希算法根据系统当前的状态以及PG的ID号，将各个PG分发到OSD集群中。

**五、Ceph的优势：**   
1.Ceph的核心RADOS通常是由少量的负责集群管理的Monitor进程和大量的负责数据存储的OSD进程构成，采用无中心节点的分布式架构，对数据进行分块多份存储。具有良好的扩展性和高可用性。   
2. Ceph分布式文件系统提供了多种客户端，包括对象存储接口、块存储接口以及文件系统接口，具有广泛的适用性，并且客户端与存储数据的OSD设备直接进行数据交互，大大提高了数据的存取性能。   
3.Ceph作为分布式文件系统，其能够在维护 POSIX 兼容性的同时加入了复制和容错功能。从2010 年 3 月底，以及可以在Linux 内核（从2.6.34版开始）中找到 Ceph 的身影，作为Linux的文件系统备选之一，Ceph.ko已经集成入Linux内核之中。Ceph 不仅仅是一个文件系统，还是一个有企业级功能的对象存储生态环境。

**六、案例：搭建Ceph分布式存储；**

**案例环境：**

|  |  |  |  |
| --- | --- | --- | --- |
| 系统 | IP地址 | 主机名（登录用户） | 承载角色 |
| Centos 7.4 64Bit 1708 | 192.168.100.101 | dlp（dhhy） | admin-node |
| Centos 7.4 64Bit 1708 | 192.168.100.102 | node1（dhhy）  添加一块硬盘 | mon-node  osd0-node  mds-node |
| Centos 7.4 64Bit 1708 | 192.168.100.103 | node2（dhhy）  添加一块硬盘 | mon-node  osd1-node |
| Centos 7.4 64Bit 1708 | 192.168.100.104 | ceph-client（root） | ceph-client |

**案例步骤：**

* 配置基础环境：
* 配置ntp时间服务；
* 分别在dlp节点、node1、node2节点、client客户端节点上安装Ceph程序；
* 在dlp节点管理node存储节点，安装注册服务，节点信息；
* 配置Ceph的mon监控进程；
* 配置Ceph的osd存储进程；
* 验证查看ceph集群状态信息：
* 配置Ceph的mds元数据进程；
* 配置Ceph的client客户端；
* 测试Ceph的客户端存储；
* 错误整理；
* **配置基础环境：**

[root@dlp ~]# useradd dhhy

[root@dlp ~]# echo "dhhy" |passwd --stdin dhhy

[root@dlp ~]# cat <<END >>/etc/hosts

192.168.100.101 dlp

192.168.100.102 node1

192.168.100.103 node2

192.168.100.104 ceph-client

END

[root@dlp ~]# echo "dhhy ALL = (root) NOPASSWD:ALL" >> /etc/sudoers.d/dhhy

[root@dlp ~]# chmod 0440 /etc/sudoers.d/dhhy

[root@node1~]# useradd dhhy

[root@node1 ~]# echo "dhhy" |passwd --stdin dhhy

[root@node1 ~]# cat <<END >>/etc/hosts

192.168.100.101 dlp

192.168.100.102 node1

192.168.100.103 node2

192.168.100.104 ceph-client

END

[root@node1 ~]# echo "dhhy ALL = (root) NOPASSWD:ALL" >> /etc/sudoers.d/dhhy

[root@node1 ~]# chmod 0440 /etc/sudoers.d/dhhy

[root@node2 ~]# useradd dhhy

[root@node2 ~]# echo "dhhy" |passwd --stdin dhhy

[root@node2 ~]# cat <<END >>/etc/hosts

192.168.100.101 dlp

192.168.100.102 node1

192.168.100.103 node2

192.168.100.104 ceph-client

END

[root@node2 ~]# echo "dhhy ALL = (root) NOPASSWD:ALL" >> /etc/sudoers.d/dhhy

[root@node2 ~]# chmod 0440 /etc/sudoers.d/dhhy

[root@ceph-client ~]# useradd dhhy

[root@ceph-client ~]# echo "dhhy" |passwd --stdin dhhy

[root@ceph-client ~]# cat <<END >>/etc/hosts

192.168.100.101 dlp

192.168.100.102 node1

192.168.100.103 node2

192.168.100.104 ceph-client

END

[root@ceph-client ~]# echo "dhhy ALL = (root) NOPASSWD:ALL" >> /etc/sudoers.d/dhhy

[root@ceph-client ~]# chmod 0440 /etc/sudoers.d/dhhy

* **配置ntp时间服务；**

[root@dlp ~]# yum -y install ntp ntpdate

[root@dlp ~]# sed -i '/^server/s/^/#/g' /etc/ntp.conf

[root@dlp ~]# sed -i '25aserver 127.127.1.0\nfudge 127.127.1.0 stratum 8' /etc/ntp.conf

[root@dlp ~]# systemctl start ntpd

[root@dlp ~]# systemctl enable ntpd

[root@dlp ~]# netstat -utpln

[root@node1 ~]# yum -y install ntpdate

[root@node1 ~]# /usr/sbin/ntpdate 192.168.100.101

[root@node1 ~]# echo "/usr/sbin/ntpdate 192.168.100.101" >>/etc/rc.local

[root@node1 ~]# chmod +x /etc/rc.local

[root@node2 ~]# yum -y install ntpdate

[root@node2 ~]# /usr/sbin/ntpdate 192.168.100.101

[root@node1 ~]# echo "/usr/sbin/ntpdate 192.168.100.101" >>/etc/rc.local

[root@node1 ~]# chmod +x /etc/rc.local

[root@ceph-client ~]# yum -y install ntpdate

[root@ceph-client ~]# /usr/sbin/ntpdate 192.168.100.101

[root@ceph-client ~]# echo "/usr/sbin/ntpdate 192.168.100.101" >>/etc/rc.local

[root@ceph-client ~]# chmod +x /etc/rc.local

* **分别在dlp节点、node1、node2节点、client客户端节点上安装Ceph；**

[root@dlp ~]# yum -y install yum-utils

[root@dlp ~]# yum-config-manager --add-repo https://dl.fedoraproject.org/pub/epel/7/x86\_64/

[root@dlp ~]# yum -y install epel-release --nogpgcheck

[root@dlp ~]# cat <<END >>/etc/yum.repos.d/ceph.repo

[Ceph]

name=Ceph packages for \$basearch

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/\$basearch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[Ceph-noarch]

name=Ceph noarch packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/noarch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[ceph-source]

name=Ceph source packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/SRPMS

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

END

[root@dlp ~]# ls /etc/yum.repos.d/ ####必须保证有默认的官网源，结合epel源和网易的ceph源，才可以进行安装；

bak CentOS-fasttrack.repo ceph.repo

CentOS-Base.repo CentOS-Media.repo dl.fedoraproject.org\_pub\_epel\_7\_x86\_64\_.repo

CentOS-CR.repo CentOS-Sources.repo epel.repo

CentOS-Debuginfo.repo CentOS-Vault.repo epel-testing.repo  
[root@dlp ~]# su - dhhy

[dhhy@dlp ~]$ mkdir ceph-cluster ##创建ceph主目录

[dhhy@dlp ~]$ cd ceph-cluster

[dhhy@dlp ceph-cluster]$ sudo yum -y install ceph-deploy ##安装ceph管理工具

[dhhy@dlp ceph-cluster]$ sudo yum -y install ceph --nogpgcheck ##安装ceph主程序

[root@node1 ~]# yum -y install yum-utils

[root@ node1 ~]# yum-config-manager --add-repo https://dl.fedoraproject.org/pub/epel/7/x86\_64/

[root@node1 ~]# yum -y install epel-release --nogpgcheck

[root@node1 ~]# cat <<END >>/etc/yum.repos.d/ceph.repo

[Ceph]

name=Ceph packages for \$basearch

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/\$basearch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[Ceph-noarch]

name=Ceph noarch packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/noarch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[ceph-source]

name=Ceph source packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/SRPMS

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

END

[root@node1 ~]# ls /etc/yum.repos.d/ ####必须保证有默认的官网源，结合epel源和网易的ceph源，才可以进行安装；

bak CentOS-fasttrack.repo ceph.repo

CentOS-Base.repo CentOS-Media.repo dl.fedoraproject.org\_pub\_epel\_7\_x86\_64\_.repo

CentOS-CR.repo CentOS-Sources.repo epel.repo

CentOS-Debuginfo.repo CentOS-Vault.repo epel-testing.repo  
[root@node1 ~]# su - dhhy

[dhhy@node1 ~]$ mkdir ceph-cluster

[dhhy@node1~]$ cd ceph-cluster

[dhhy@node1 ceph-cluster]$ sudo yum -y install ceph-deploy

[dhhy@node1 ceph-cluster]$ sudo yum -y install ceph --nogpgcheck

[dhhy@node1 ceph-cluster]$ sudo yum -y install deltarpm

[root@node2 ~]# yum -y install yum-utils

[root@ node1 ~]# yum-config-manager --add-repo https://dl.fedoraproject.org/pub/epel/7/x86\_64/

[root@node2 ~]# yum -y install epel-release --nogpgcheck

[root@node2 ~]# cat <<END >>/etc/yum.repos.d/ceph.repo

[Ceph]

name=Ceph packages for \$basearch

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/\$basearch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[Ceph-noarch]

name=Ceph noarch packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/noarch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[ceph-source]

name=Ceph source packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/SRPMS

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

END

[root@node2 ~]# ls /etc/yum.repos.d/ ####必须保证有默认的官网源，结合epel源和网易的ceph源，才可以进行安装；

bak CentOS-fasttrack.repo ceph.repo

CentOS-Base.repo CentOS-Media.repo dl.fedoraproject.org\_pub\_epel\_7\_x86\_64\_.repo

CentOS-CR.repo CentOS-Sources.repo epel.repo

CentOS-Debuginfo.repo CentOS-Vault.repo epel-testing.repo  
[root@node2 ~]# su - dhhy

[dhhy@node2 ~]$ mkdir ceph-cluster

[dhhy@node2 ~]$ cd ceph-cluster

[dhhy@node2 ceph-cluster]$ sudo yum -y install ceph-deploy

[dhhy@node2 ceph-cluster]$ sudo yum -y install ceph --nogpgcheck

[dhhy@node2 ceph-cluster]$ sudo yum -y install deltarpm

[root@ceph-client ~]# yum -y install yum-utils

[root@ node1 ~]# yum-config-manager --add-repo https://dl.fedoraproject.org/pub/epel/7/x86\_64/

[root@ceph-client ~]# yum -y install epel-release --nogpgcheck

[root@ceph-client ~]# cat <<END >>/etc/yum.repos.d/ceph.repo

[Ceph]

name=Ceph packages for \$basearch

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/\$basearch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[Ceph-noarch]

name=Ceph noarch packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/noarch

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

[ceph-source]

name=Ceph source packages

baseurl=http://mirrors.163.com/ceph/rpm-jewel/el7/SRPMS

enabled=1

gpgcheck=0

type=rpm-md

gpgkey=https://mirrors.163.com/ceph/keys/release.asc

priority=1

END

[root@ceph-client ~]# ls /etc/yum.repos.d/ ####必须保证有默认的官网源，结合epel源和网易的ceph源，才可以进行安装；

bak CentOS-fasttrack.repo ceph.repo

CentOS-Base.repo CentOS-Media.repo dl.fedoraproject.org\_pub\_epel\_7\_x86\_64\_.repo

CentOS-CR.repo CentOS-Sources.repo epel.repo

CentOS-Debuginfo.repo CentOS-Vault.repo epel-testing.repo

[root@ceph-client ~]# yum -y install yum-plugin-priorities

[root@ceph-client ~]# yum -y install ceph ceph-radosgw --nogpgcheck

* **在dlp节点管理node存储节点，安装注册服务，节点信息；**

[dhhy@dlp ceph-cluster]$ pwd ##当前目录必须为ceph的安装目录位置

/home/dhhy/ceph-cluster

[dhhy@dlp ceph-cluster]$ ssh-keygen -t rsa ##主节点需要远程管理mon节点，需要创建密钥对，并且将公钥复制到mon节点

[dhhy@dlp ceph-cluster]$ ssh-copy-id dhhy@dlp

[dhhy@dlp ceph-cluster]$ ssh-copy-id dhhy@node1

[dhhy@dlp ceph-cluster]$ ssh-copy-id dhhy@node2

[dhhy@dlp ceph-cluster]$ ssh-copy-id root@ceph-client

[dhhy@dlp ceph-cluster]$ cat <<END >>/home/dhhy/.ssh/config

Host dlp

Hostname dlp

User dhhy

Host node1

Hostname node1

User dhhy

Host node2

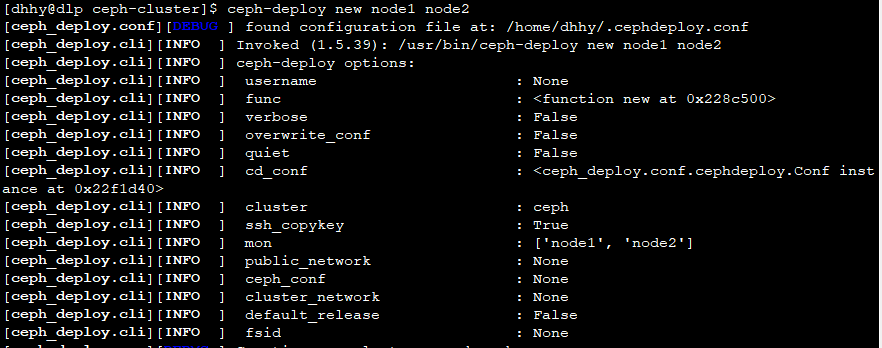
Hostname node2

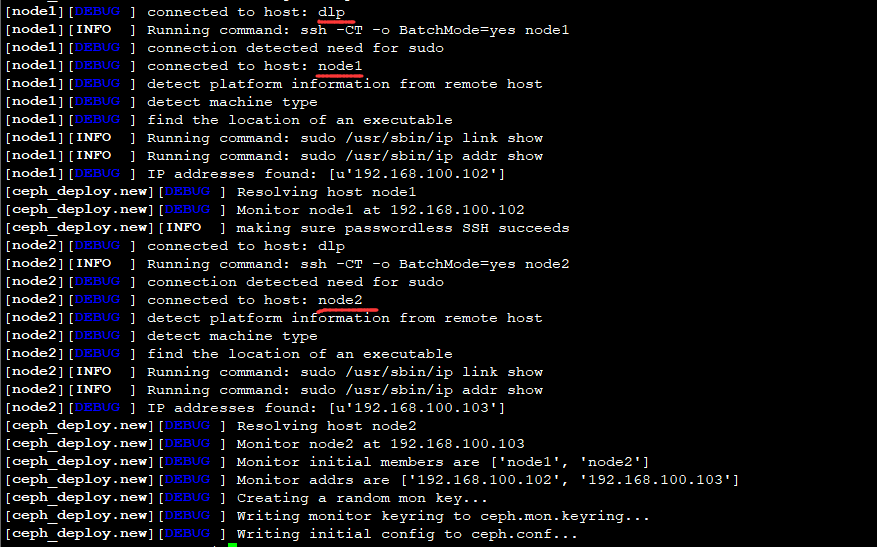
User dhhy

END

[dhhy@dlp ceph-cluster]$ chmod 644 /home/dhhy/.ssh/config

[dhhy@dlp ceph-cluster]$ ceph-deploy new node1 node2 ##初始化节点





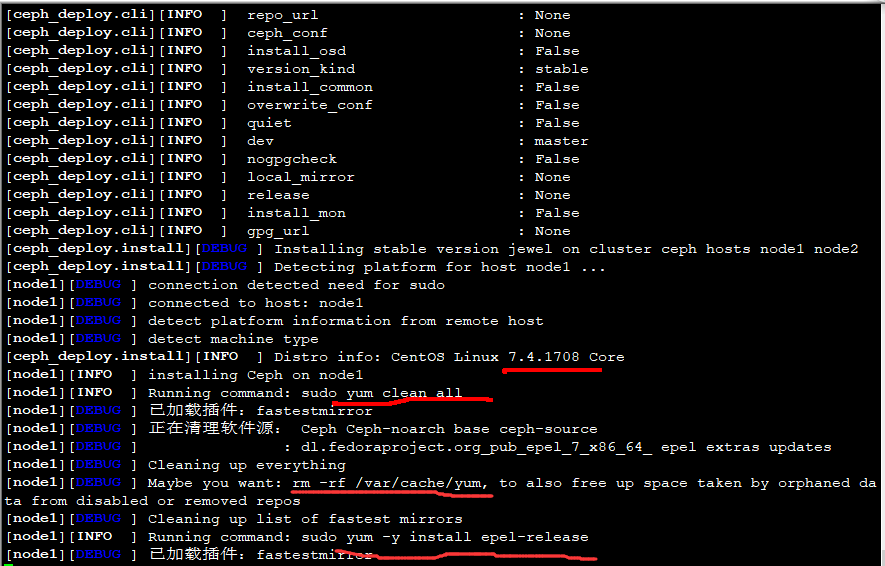


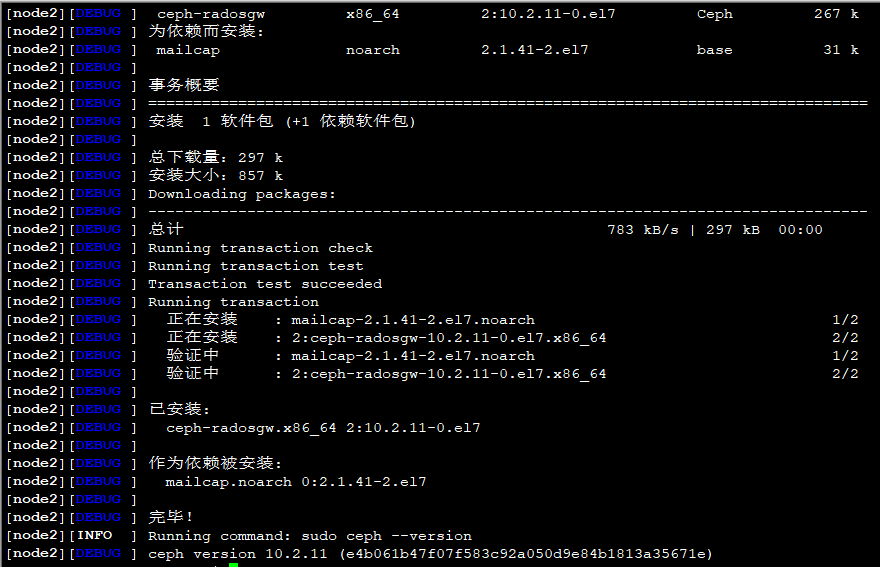
[dhhy@dlp ceph-cluster]$ cat <<END >>/home/dhhy/ceph-cluster/ceph.conf

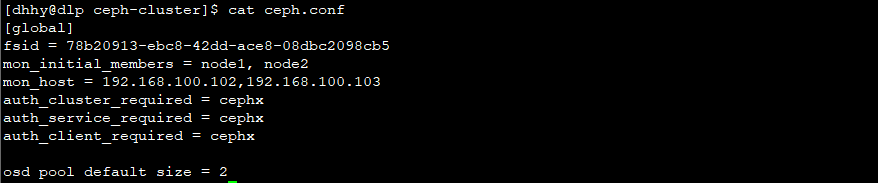
osd pool default size = 2

END

[dhhy@dlp ceph-cluster]$ ceph-deploy install node1 node2 ##安装ceph

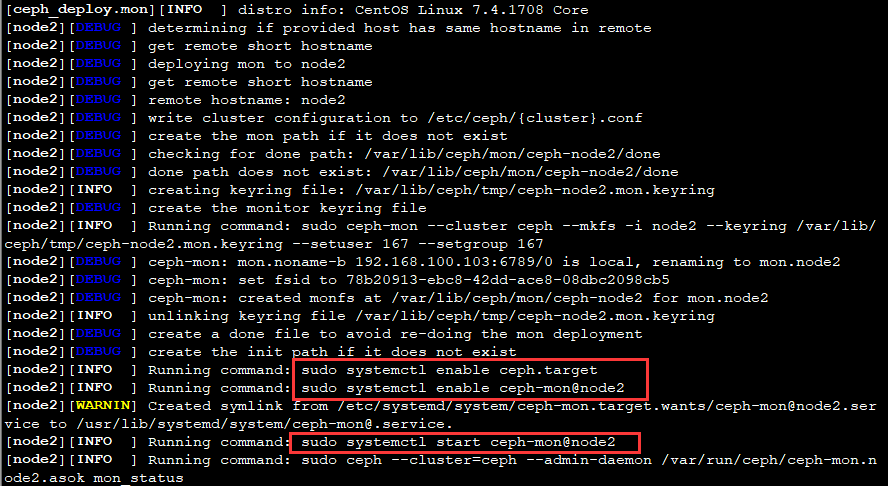


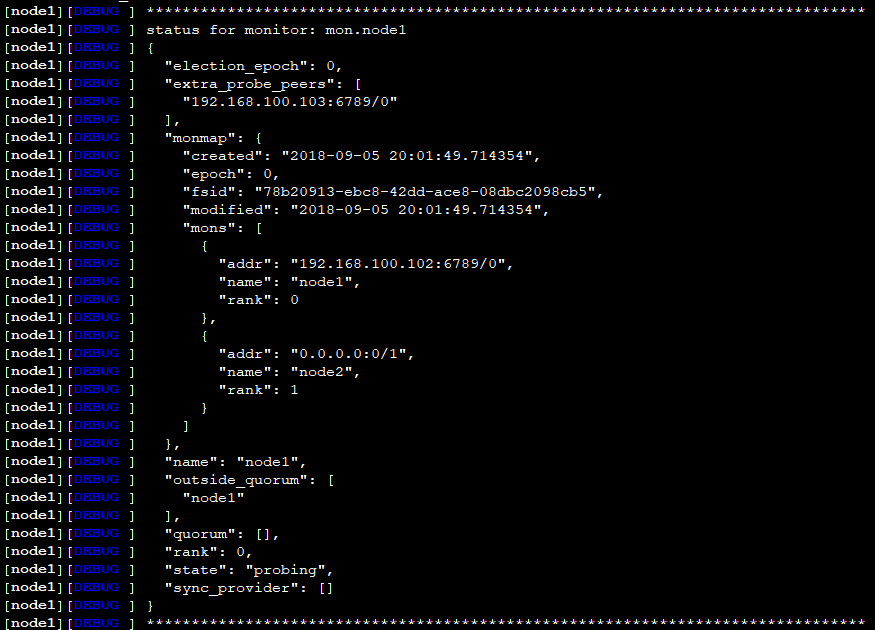


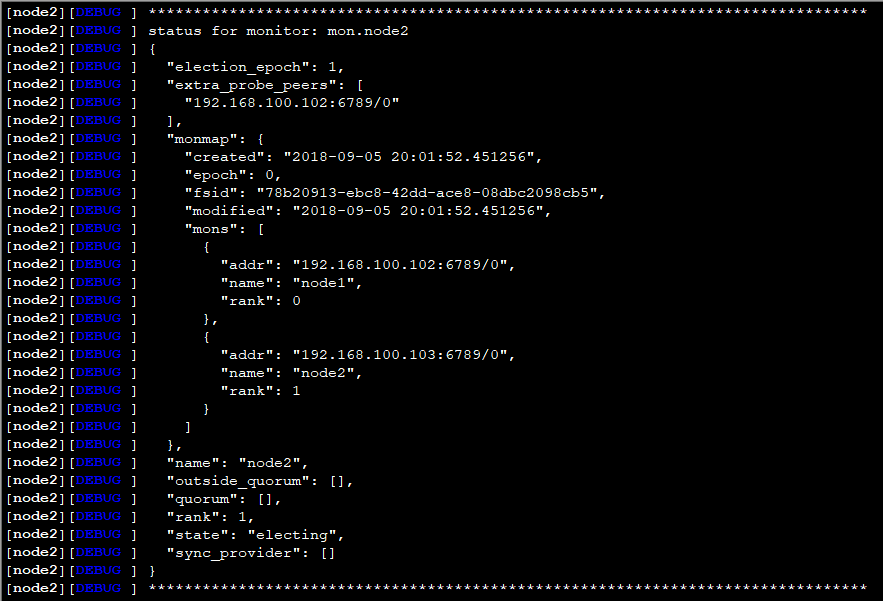


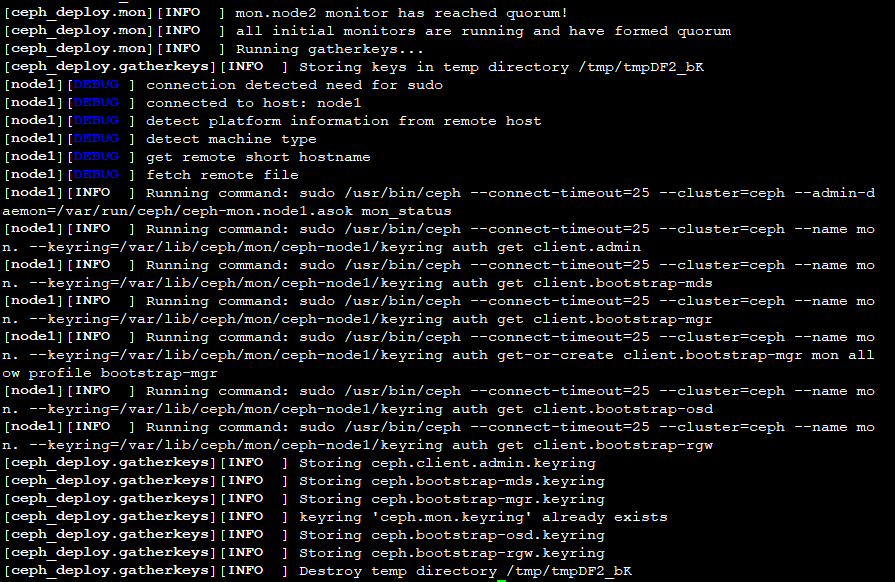
* **配置Ceph的mon监控进程；**

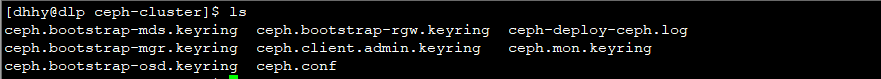
[dhhy@dlp ceph-cluster]$ ceph-deploy mon create-initial ##初始化mon节点

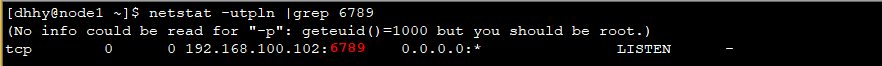


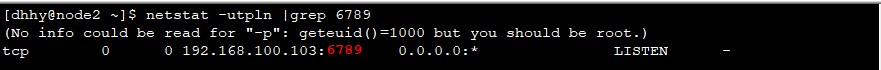












注解：node节点的配置文件在/etc/ceph/目录下，会自动同步dlp管理节点的配置文件；

* **配置Ceph的osd存储；**

**配置node1节点的osd0存储设备：**

[dhhy@dlp ceph-cluster]$ ssh dhhy@node1 ##创建osd节点存储数据的目录位置

[dhhy@node1 ~]$ sudo fdisk /dev/sdb

n p 回车 回车 回车 p w

[dhhy@node1 ~]$ sudo partx -a /dev/sdb

[dhhy@node1 ~]$ sudo mkfs -t xfs /dev/sdb1

[dhhy@node1 ~]$ sudo mkdir /var/local/osd0

[dhhy@node1 ~]$ sudo vi /etc/fstab

/dev/sdb1 /var/local/osd0 xfs defaults 0 0

:wq

[dhhy@node1 ~]$ sudo mount -a

[dhhy@node1 ~]$ sudo chmod 777 /var/local/osd0

[dhhy@node1 ~]$ sudo chown ceph:ceph /var/local/osd0/

[dhhy@node1 ~]$ ls -ld /var/local/osd0/

[dhhy@node1 ~]$ df -hT

[dhhy@node1 ~]$ exit

**配置node2节点的osd1存储设备：**

[dhhy@dlp ceph-cluster]$ ssh dhhy@node2

[dhhy@node2 ~]$ sudo fdisk /dev/sdb

n p 回车 回车 回车 p w

[dhhy@node2 ~]$ sudo partx -a /dev/sdb

[dhhy@node2 ~]$ sudo mkfs -t xfs /dev/sdb1

[dhhy@node2 ~]$ sudo mkdir /var/local/osd1

[dhhy@node2 ~]$ sudo vi /etc/fstab

/dev/sdb1 /var/local/osd1 xfs defaults 0 0

:wq

[dhhy@node2 ~]$ sudo mount -a

[dhhy@node2 ~]$ sudo chmod 777 /var/local/osd1

[dhhy@node2 ~]$ sudo chown ceph:ceph /var/local/osd1/

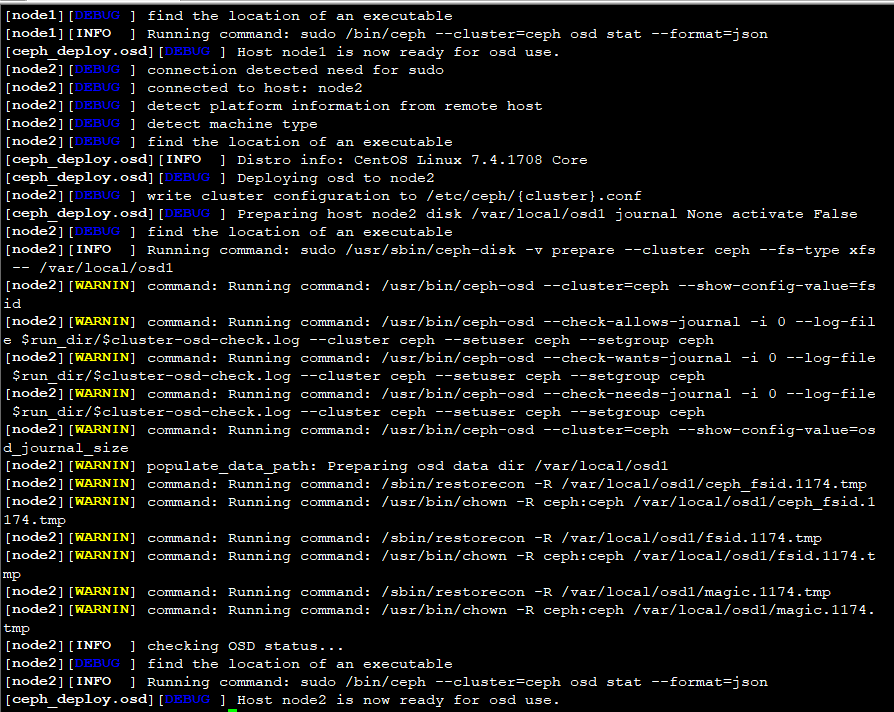
[dhhy@node2~]$ ls -ld /var/local/osd1/

[dhhy@node2 ~]$ df -hT

[dhhy@node2 ~]$ exit

**dlp管理节点注册node节点：**

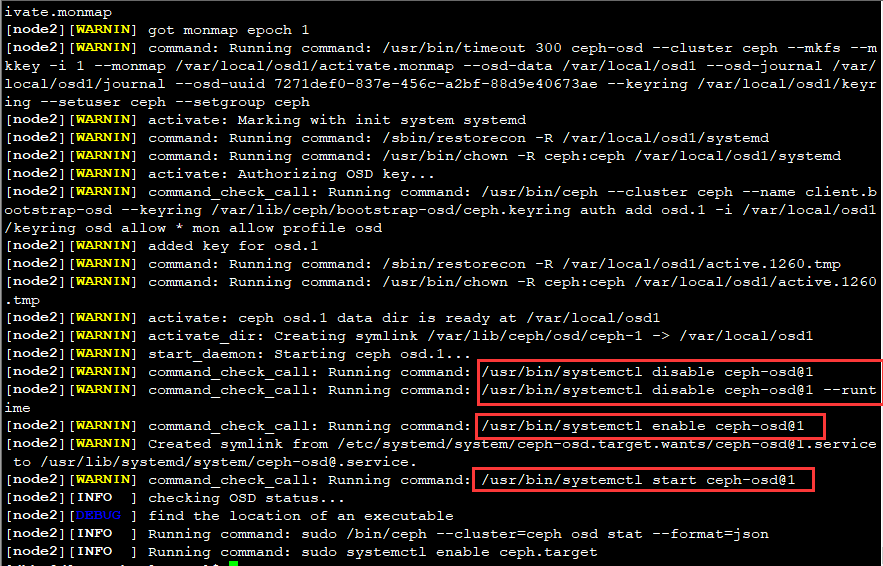
[dhhy@dlp ceph-cluster]$ ceph-deploy osd prepare node1:/var/local/osd0 node2:/var/local/osd1 ##初始创建osd节点并指定节点存储文件位置



[dhhy@dlp ceph-cluster]$ chmod +r /home/dhhy/ceph-cluster/ceph.client.admin.keyring

[dhhy@dlp ceph-cluster]$ ceph-deploy osd activate node1:/var/local/osd0 node2:/var/local/osd1

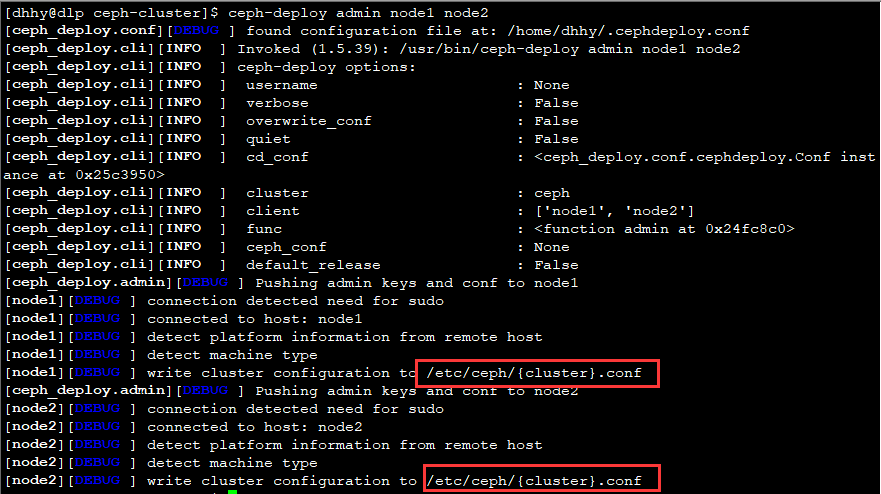
##激活ods节点







[dhhy@dlp ceph-cluster]$ ceph-deploy admin node1 node2 ##复制key管理密钥文件到node节点中



[dhhy@dlp ceph-cluster]$ sudo cp /home/dhhy/ceph-cluster/ceph.client.admin.keyring /etc/ceph/

[dhhy@dlp ceph-cluster]$ sudo cp /home/dhhy/ceph-cluster/ceph.conf /etc/ceph/

[dhhy@dlp ceph-cluster]$ ls /etc/ceph/

ceph.client.admin.keyring ceph.conf rbdmap

[dhhy@dlp ceph-cluster]$ ceph quorum\_status --format json-pretty ##查看Ceph群集详细信息

{

"election\_epoch": 4,

"quorum": [

0,

1

],

"quorum\_names": [

"node1",

"node2"

],

"quorum\_leader\_name": "node1",

"monmap": {

"epoch": 1,

"fsid": "dc679c6e-29f5-4188-8b60-e9eada80d677",

"modified": "2018-06-02 23:54:34.033254",

"created": "2018-06-02 23:54:34.033254",

"mons": [

{

"rank": 0,

"name": "node1",

"addr": "192.168.100.102:6789\/0"

},

{

"rank": 1,

"name": "node2",

"addr": "192.168.100.103:6789\/0"

}

]

}

}

* **验证查看ceph集群状态信息：**

[dhhy@dlp ceph-cluster]$ ceph health

HEALTH\_OK

[dhhy@dlp ceph-cluster]$ ceph -s ##查看Ceph群集状态

cluster 24fb6518-8539-4058-9c8e-d64e43b8f2e2

health HEALTH\_OK

monmap e1: 2 mons at {node1=192.168.100.102:6789/0,node2=192.168.100.103:6789/0}

election epoch 6, quorum 0,1 node1,node2

osdmap e10: 2 osds: 2 up, 2 in

flags sortbitwise,require\_jewel\_osds

pgmap v20: 64 pgs, 1 pools, 0 bytes data, 0 objects

10305 MB used, 30632 MB / 40938 MB avail ##已使用、剩余、总容量

64 active+clean

[dhhy@dlp ceph-cluster]$ ceph osd tree

ID WEIGHT TYPE NAME UP/DOWN REWEIGHT PRIMARY-AFFINITY

-1 0.03897 root default

-2 0.01949 host node1

0 0.01949 osd.0 up 1.00000 1.00000

-3 0.01949 host node2

1 0.01949 osd.1 up 1.00000 1.00000

[dhhy@dlp ceph-cluster]$ ssh dhhy@node1 ##验证node1节点的端口监听状态以及其配置文件以及磁盘使用情况

[dhhy@node1 ~]$ df -hT |grep sdb1

/dev/sdb1 xfs 20G 5.1G 15G 26% /var/local/osd0

[dhhy@node1 ~]$ du -sh /var/local/osd0/

5.1G /var/local/osd0/

[dhhy@node1 ~]$ ls /var/local/osd0/

activate.monmap active ceph\_fsid current fsid journal keyring magic ready store\_version superblock systemd type whoami

[dhhy@node1 ~]$ ls /etc/ceph/

ceph.client.admin.keyring ceph.conf rbdmap tmppVBe\_2

[dhhy@node1 ~]$ cat /etc/ceph/ceph.conf

[global]

fsid = 0fcdfa46-c8b7-43fc-8105-1733bce3bfeb

mon\_initial\_members = node1, node2

mon\_host = 192.168.100.102,192.168.100.103

auth\_cluster\_required = cephx

auth\_service\_required = cephx

auth\_client\_required = cephx

osd pool default size = 2

[dhhy@node1 ~]$ exit

[dhhy@dlp ceph-cluster]$ ssh dhhy@node2 ##验证node2节点的端口监听状态以及其配置文件及其磁盘使用情况

[dhhy@node2 ~]$ df -hT |grep sdb1

/dev/sdb1 xfs 20G 5.1G 15G 26% /var/local/osd1

[dhhy@node2 ~]$ du -sh /var/local/osd1/

5.1G /var/local/osd1/

[dhhy@node2 ~]$ ls /var/local/osd1/

activate.monmap active ceph\_fsid current fsid journal keyring magic ready store\_version superblock systemd type whoami

[dhhy@node2 ~]$ ls /etc/ceph/

ceph.client.admin.keyring ceph.conf rbdmap tmpmB\_BTa

[dhhy@node2 ~]$ cat /etc/ceph/ceph.conf

[global]

fsid = 0fcdfa46-c8b7-43fc-8105-1733bce3bfeb

mon\_initial\_members = node1, node2

mon\_host = 192.168.100.102,192.168.100.103

auth\_cluster\_required = cephx

auth\_service\_required = cephx

auth\_client\_required = cephx

osd pool default size = 2

[dhhy@node2 ~]$ exit

* **配置Ceph的mds元数据进程；**

[dhhy@dlp ceph-cluster]$ ceph-deploy mds create node1

[dhhy@dlp ceph-cluster]$ ssh dhhy@node1

[dhhy@node1 ~]$ netstat -utpln |grep 68

(No info could be read for "-p": geteuid()=1000 but you should be root.)

tcp 0 0 0.0.0.0:6800 0.0.0.0:\* LISTEN -

tcp 0 0 0.0.0.0:6801 0.0.0.0:\* LISTEN -

tcp 0 0 0.0.0.0:6802 0.0.0.0:\* LISTEN -

tcp 0 0 0.0.0.0:6803 0.0.0.0:\* LISTEN -

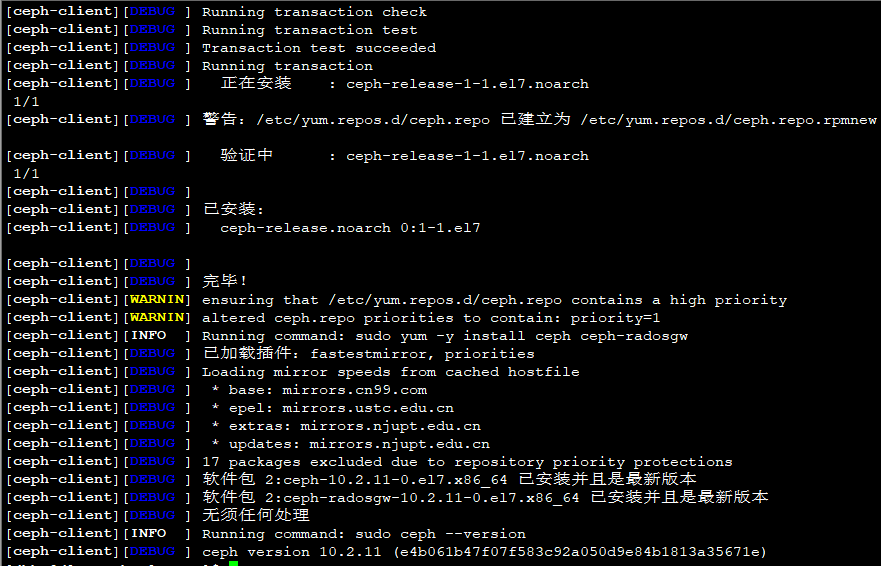
tcp 0 0 0.0.0.0:6804 0.0.0.0:\* LISTEN -

tcp 0 0 192.168.100.102:6789 0.0.0.0:\* LISTEN -

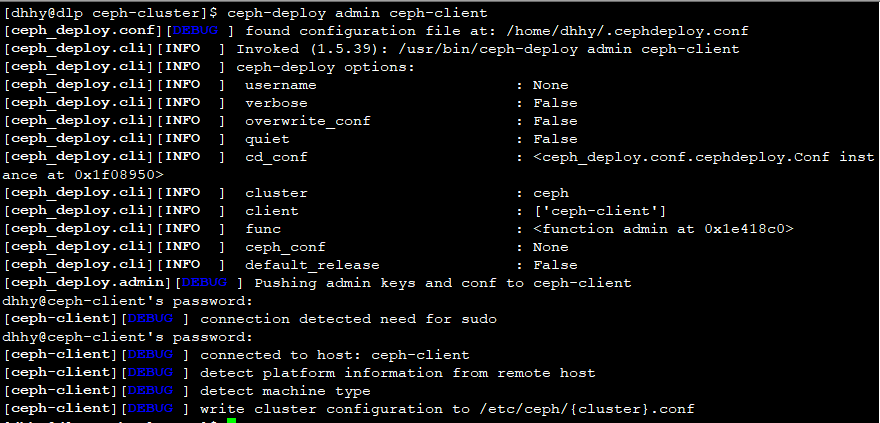
[dhhy@node1 ~]$ exit

* **配置Ceph的client客户端；**

[dhhy@dlp ceph-cluster]$ ceph-deploy install ceph-client ##提示输入密码，请输入dhhy，温柔点；



[dhhy@dlp ceph-cluster]$ ceph-deploy admin ceph-client



[dhhy@dlp ceph-cluster]$ ssh root@ceph-client

[root@ceph-client ~]# chmod +r /etc/ceph/ceph.client.admin.keyring

[root@ceph-client ~]# exit

[dhhy@dlp ceph-cluster]$ ceph osd pool create cephfs\_data 128 ##数据存储池

pool 'cephfs\_data' created

[dhhy@dlp ceph-cluster]$ ceph osd pool create cephfs\_metadata 128 ##元数据存储池

pool 'cephfs\_metadata' created

[dhhy@dlp ceph-cluster]$ ceph fs new cephfs cephfs\_data cephfs\_metadata ##创建文件系统

new fs with metadata pool 1 and data pool 2

[dhhy@dlp ceph-cluster]$ ceph fs ls ##查看文件系统

name: cephfs, metadata pool: cephfs\_data, data pools: [cephfs\_metadata ]

[dhhy@dlp ceph-cluster]$ ceph -s

cluster 24fb6518-8539-4058-9c8e-d64e43b8f2e2

health HEALTH\_WARN

clock skew detected on mon.node2

too many PGs per OSD (320 > max 300)

Monitor clock skew detected

monmap e1: 2 mons at {node1=192.168.100.102:6789/0,node2=192.168.100.103:6789/0}

election epoch 6, quorum 0,1 node1,node2

fsmap e5: 1/1/1 up {0=node1=up:active}

osdmap e17: 2 osds: 2 up, 2 in

flags sortbitwise,require\_jewel\_osds

pgmap v54: 320 pgs, 3 pools, 4678 bytes data, 24 objects

10309 MB used, 30628 MB / 40938 MB avail

320 active+clean

* **测试Ceph的客户端存储；**

[dhhy@dlp ceph-cluster]$ ssh root@ceph-client

[root@ceph-client ~]# mkdir /mnt/ceph

[root@ceph-client ~]# grep key /etc/ceph/ceph.client.admin.keyring |awk '{print $3}' >>/etc/ceph/admin.secret

[root@ceph-client ~]# cat /etc/ceph/admin.secret

AQCd/x9bsMqKFBAAZRNXpU5QstsPlfe1/FvPtQ==

[root@ceph-client ~]# mount -t ceph 192.168.100.102:6789:/ /mnt/ceph/ -o name=admin,secretfile=/etc/ceph/admin.secret

[root@ceph-client ~]# df -hT |grep ceph

192.168.100.102:6789:/ ceph 40G 11G 30G 26% /mnt/ceph

[root@ceph-client ~]# dd if=/dev/zero of=/mnt/ceph/1.file bs=1G count=1

记录了1+0 的读入

记录了1+0 的写出

1073741824字节(1.1 GB)已复制，14.2938 秒，75.1 MB/秒

[root@ceph-client ~]# ls /mnt/ceph/

1.file  
[root@ceph-client ~]# df -hT |grep ceph

192.168.100.102:6789:/ ceph 40G 13G 28G 33% /mnt/ceph

[root@ceph-client ~]# mkdir /mnt/ceph1

[root@ceph-client ~]# mount -t ceph 192.168.100.103:6789:/ /mnt/ceph1/ -o name=admin,secretfile=/etc/ceph/admin.secret

[root@ceph-client ~]# df -hT |grep ceph

192.168.100.102:6789:/ ceph 40G 15G 26G 36% /mnt/ceph

192.168.100.103:6789:/ ceph 40G 15G 26G 36% /mnt/ceph1

[root@ceph-client ~]# ls /mnt/ceph1/

1.file 2.file

* **错误整理：**

1. 如若在配置过程中出现问题，重新创建集群或重新安装ceph，那么需要将ceph集群中的数据都清除掉，命令如下；

[dhhy@dlp ceph-cluster]$ ceph-deploy purge node1 node2

[dhhy@dlp ceph-cluster]$ ceph-deploy purgedata node1 node2

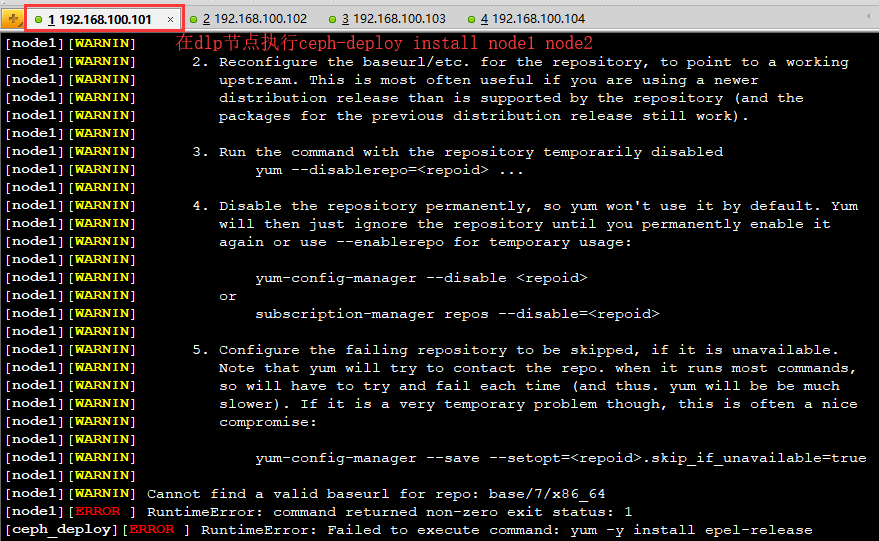
[dhhy@dlp ceph-cluster]$ ceph-deploy forgetkeys && rm ceph.\*

2.dlp节点为node节点和客户端安装ceph时，会出现yum安装超时，大多由于网络问题导致，可以多执行几次安装命令；

3.dlp节点指定ceph-deploy命令管理node节点配置时，当前所在目录一定是/home/dhhy/ceph-cluster/，不然会提示找不到ceph.conf的配置文件；

4.osd节点的/var/local/osd\*/存储数据实体的目录权限必须为777，并且属主和属组必须为ceph；

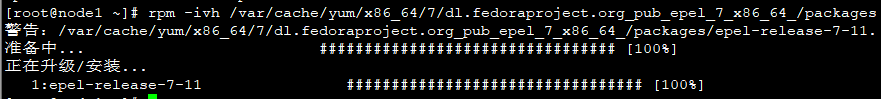
5. 在dlp管理节点安装ceph时出现以下问题



解决方法：

1.重新yum安装node1或者node2的epel-release软件包；

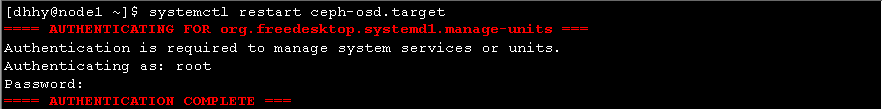
2.如若还无法解决，将软件包下载，使用以下命令进行本地安装；

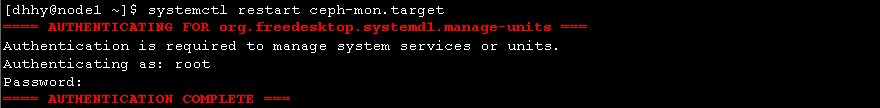


**6.如若在dlp管理节点中对/home/dhhy/ceph-cluster/ceph.conf主配置文件发生变化，那么需要将其主配置文件同步给node节点，命令如下：**

****

**node节点收到配置文件后，需要重新启动进程：**

****

****

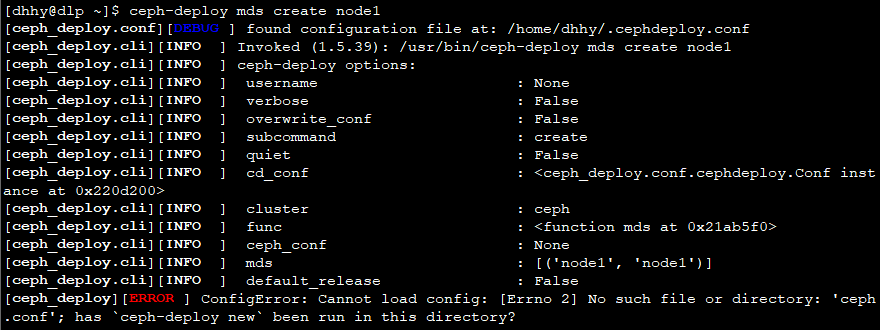
**7.在dlp管理节点查看ceph集群状态时，出现如下，原因是因为时间不一致所导致；**

****

**解决方法：将dlp节点的ntpd时间服务重新启动，node节点再次同步时间即可，如下所示：**

****

**8.在dlp管理节点进行管理node节点时，所处的位置一定是/home/dhhy/ceph-cluster/，不然会提示找不到ceph.conf主配置文件；**

****