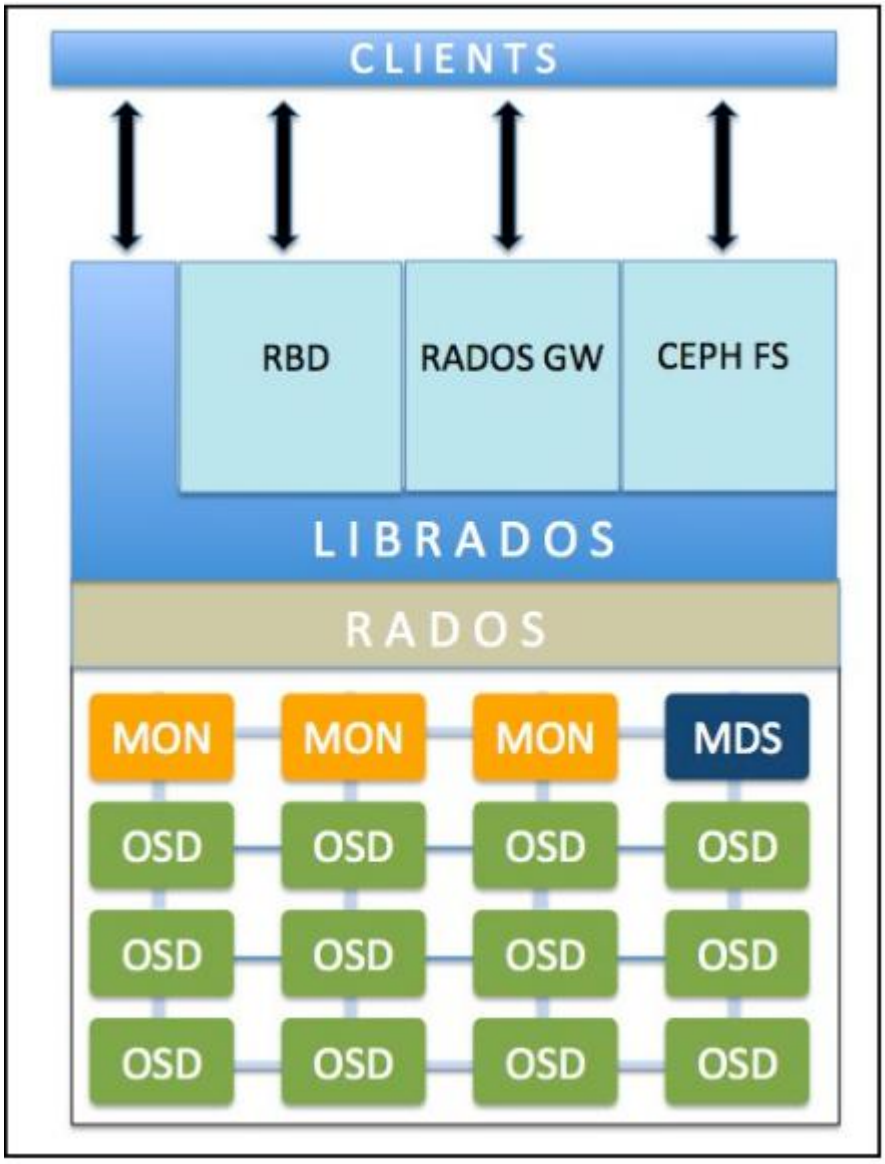
存储：
RAID：独立磁盘冗余阵列
性能、容错、空间

分布式存储：CEPH
1、组成



（1）MON：监视器。MON 通过保存一系列集群状态 map 来监视集群的组件。MON 因为保存集群状态，要防止单点故障，所以需要多台；另外，MON 需要是奇数，如果出现意见分岐，采用投票机制，少数服从多数。
（2）OSD：对象存储设备。真正存储数据的组件。一般来说，每块参与存储的磁盘都需要一个 OSD 进程。
（3）MDS：元数据服务器。只有 CephFS 需要它。
元数据：metadata，存储数据的数据。比如一本书内容是数据，那么书的作者、出版社、出版时间之类的信息就是元数据。
（4）RADOS：可靠自主分 ot@room8pc16 nsd2018]# for i in {1..6}

（5）> do
（6）> echo -e "192.168.4.$i\tnode$i.tedu.cn\tnode$i" >> /etc/hosts
（7）> done
（8）布式对象存储。它是 ceph 存储的基础，保证一切都以对象形式存储。
（9）RBD：RADOS 块设备，提供块存储
（10）CephFS：提供文件系统级别存储
（11）RGW：RADOS 网关，提供对象存储

存储分类：
块存储：提供硬盘，如 iSCSI
文件级别存储：共享文件夹
对象存储：一切皆对象
http://storage.ctocio.com.cn/281/12110781_2.shtml

CEPH 环境准备
1、准备 6 台虚拟机
主机名、IP 地址
2、在物理主机上配置名称解析
[root@room8pc16 nsd2018]# for i in {1..6}
> do
> echo -e "192.168.4.$i\tnode$i.tedu.cn\tnode$i" >> /etc/hosts
> done
执行结果如下：
[root@room8pc16 nsd2018]#
。。。。
192.168.4.1 node1.tedu.cn node1
192.168.4.2 node2.tedu.cn node2
192.168.4.3 node3.tedu.cn node3
192.168.4.4 node4.tedu.cn node4
192.168.4.5 node5.tedu.cn node5
192.168.4.6 node6.tedu.cn node6

\t -->tab 键
3、提前将服务器的密钥保存，不需要 ssh 时回答 yes
[root@room8pc16 nsd2018]# ssh-keyscan node{1..6} >> /root/.ssh/known_hosts
4、实现免密登陆
[root@room8pc16 nsd2018]# for i in {1..6}
> do
> ssh-copy-id node$i
> done

5、配置 yum 源

```
[root@room8pc16 nsd2018]# mkdir /var/ftp/ceph/
[root@room8pc16 nsd2018]# vim   /etc/fstab
/ISO/rhcs2.0-rhosp9-20161113-x86_64.iso  /var/ftp/ceph       iso9660
defaults      0 0
[root@room8pc16 nsd2018]# mount -a
[root@room8pc16 nsd2018]# vim /tmp/server.repo
[rhel7.4]
name=rhel7.4
baseurl=ftp://192.168.4.254/rhel7.4
enabled=1
gpgcheck=0
[mon]
name=mon
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/MON
enabled=1
gpgcheck=0
[osd]
name=osd
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/OSD
enabled=1
gpgcheck=0
[tools]
name=tools
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/Tools
enabled=1
gpgcheck=0
[root@room8pc16 nsd2018]# for vm in node{1..6}
> do
> scp /tmp/server.repo ${vm}:/etc/yum.repos.d/
> done
```

6、配置 node1 节点为管理节点

（1）配置名称解析

```
[root@node1    ~]#    for    i    in    {1..6};    do    echo    -e
"192.168.4.$i\tnode$i.tedu.cn\tnode$i" >> /etc/hosts; done
```

（2）配置免密登陆

```
[root@node1       ~]#       ssh-keyscan       node{1..6}       >>
/root/.ssh/known_hosts
[root@node1 ~]# ssh-keygen   -f   /root/.ssh/id_rsa   -N ''
[root@node1 ~]# for i in {1..6}; do ssh-copy-id node$i; done
[root@node1 ~]# for vm in node{1..6}
> do
> scp /etc/hosts ${vm}:/etc/
> done
```

7、NTP 网络时间协议，基于 UDP123 端口。用于时间同步
时区：地球一圈 360 度，经度每 15 度角一个时区，共 24 个时区。以英国格林威治这个城市所在纵切面为基准。北京在东八区。
夏季节约时间：夏令时。DST
Stratum：时间服务器的层级。
时间准确度：原子钟。

8、配置 node6 为时间服务器
（1）配置
```
[root@node6 ~]# yum install -y chrony
[root@node6 ~]# vim /etc/chrony.conf
server 0.centos.pool.ntp.org iburst
#server 1.centos.pool.ntp.org iburst
#server 2.centos.pool.ntp.org iburst
#server 3.centos.pool.ntp.org iburst
allow 192.168.4.0/24
local stratum 10
```
（2）启动服务
```
[root@node6 ~]# systemctl enable chronyd
[root@node6 ~]# systemctl restart chronyd
```

将 node1-5 配置为 NTP 的客户端
（1）配置
```
[root@node1 ~]# vim /etc/chrony.conf
#server 0.rhel.pool.ntp.org iburst
#server 1.rhel.pool.ntp.org iburst
#server 2.rhel.pool.ntp.org iburst
#server 3.rhel.pool.ntp.org iburst
server 192.168.4.6 iburst
[root@node1 ~]# systemctl restart chronyd
```
（2）测试
```
[root@node1 ~]# date -s "2018-7-13 12:00:00"
[root@node1 ~]# ntpdate 192.168.4.6
[root@node1 ~]# date
```
（3）同步其他主机
```
[root@node1 ~]# for i in {2..5}
> do
> scp /etc/chrony.conf node$i:/etc/
> done
[root@node1 ~]# for vm in node{2..5}
> do
> ssh  $vm  systemctl  restart  chronyd
> done
```
9、为 node1-3 各添加 3 块 10GB 的磁盘

可以在虚拟机不关机的情况下，直接添加硬盘

安装 ceph
1、在 node1 上安装部署软件
[root@node1 ~]# yum install -y ceph-deploy
2、创建 ceph 部署工具的工作目录
[root@node1 ~]# mkdir ceph-clu
3、创建参与集群节点的配置文件
[root@node1 ceph-clu]# ceph-deploy new node{1..3}
[root@node1 ceph-clu]# ls
4、在 3 个节点上安装软件包
[root@node1 ceph-clu]# ceph-deploy install node{1..3}
5、初始化 mon 服务
[root@node1 ceph-clu]# ceph-deploy mon create-initial
如果出现以下错误：
[node1][ERROR] admin_socket: exception getting command descriptions: [Errno 2] No such file or directory
解决方案：
[root@node1 ceph-clu]# vim ceph.conf 最下面加入行：
public_network = 192.168.4.0/24
再执行以下命令：
[root@host1 ceoh-clu]# ceph-deploy --overwrite-conf config push node1 node2 node3

6、把 node1-3 的 vdb 作为日志盘。Ext / xfs 都是日志文件系统，一个分区分成日志区和数据区。为了更好的性能，vdb 专门作为 vdc 和 vdd 的日志盘。
[root@node1 ceph-clu]# for vm in node{1..3}
> do
> ssh $vm parted /dev/vdb mklabel gpt
> done
[root@node1 ceph-clu]# for vm in node{1..3}; do ssh $vm parted /dev/vdb mkpart primary 1M 50% ; done
[root@node1 ceph-clu]# for vm in node{1..3}; do ssh $vm parted /dev/vdb mkpart primary 50% 100% ; done
[root@node1 ceph-clu]# for vm in node{1..3}; do ssh ${vm} chown ceph.ceph /dev/vdb? ; done

7、创建 OSD 设备
[root@node1 ceph-clu]# for i in {1..3}
> do
> ceph-deploy disk zap node$i:vdc node$i:vdd
> done
[root@node1 ceph-clu]# for i in {1..3}

> do
> ceph-deploy osd create node$i:vdc:/dev/vdb1 node$i:vdd:/dev/vdb2
> done
8、验证
到第 7 步为止，ceph 已经搭建完成。查看 ceph 状态
[root@node1 ceph-clu]# ceph -s 如果出现 health HEALTH_OK 表示正常

9、排错
https://www.zybuluo.com/dyj2017/note/920621

CEPH 应用
1、块存储：使用最多的一种方式
2、cephFS：了解，不建议在生产环境中使用，因为还不成熟
3、对象存储：了解，使用亚马逊的 s3

使用 RBD(Rados 块设备)
1、查看存储池
[root@node1 ~]# ceph osd lspools
可以查看到 0 号镜像池，名字为 rbd
2、创建名为 demo-img 的镜像大小为 10GB
[root@node1 ~]# rbd create demo-img --image-feature layering --size 10G
[root@node1 ~]# rbd list
[root@node1 ~]# rbd info demo-img
3、创建第 2 个镜像，名为 image，指定它位于 rbd 池中
[root@node1 ~]# rbd create rbd/image --image-feature layering --size 10G

编写 UDEV 规则，使得 vdb1 和 vdb2 重启后，属主属组仍然是 ceph
[root@node1 ~]# vim /etc/udev/rules.d/90-cephdisk.rules
ACTION=="add", KERNEL=="vdb[12]", OWNER="ceph", GROUP="ceph"

##

[root@room]# ssh node2 ls /root # ssh ip 地址 command