集群：HPC／LB／HA

LB：nginx / lvs / haproxy / F5

HA：keepalived / RHCS

LVS：Linux虚拟服务器

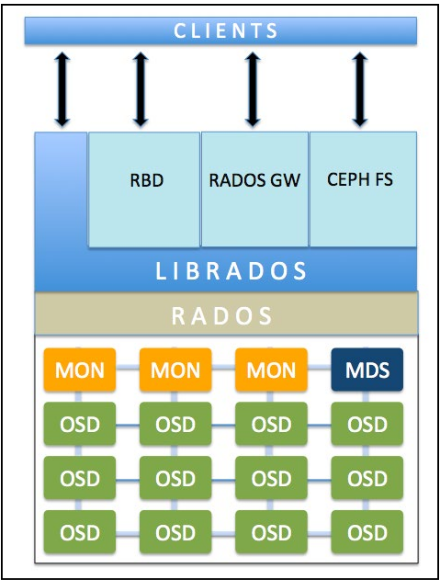
LVS工作模式：NAT ／ TUN ／ DR

存储：

1. 块存储，如iSCSI、 FC SAN
2. 文件存储，如NFS，CIFS
3. 对象存储

Ceph:

1. 它是一个分布式文件系统，可以提供块存储、文件系统存储和对象存储。但是文件存储还不非常成熟，不建议在生产环境中使用。应用的最多的形式是块存储。
2. Ceph主要组件



OSD：对象存储设备，它是ceph中唯一的真正进行数据存储的组件。通常情况下，一个OSD进程绑定到一块物理磁盘。

MON：Monitor监视器，跟踪整个集群的健康状态。它为每个ceph组件维护一个映射表。MON进程数目是奇数，如3、5、7……

MDS：元数据服务器。为ceph文件系统存储提供元数据，如果不是文件系统存储则不需要。

元数据：描述数据的数据。如一本书的出版社、页数、作者、出版时间等都是元数据。

RADOS：可靠自主的分存式对象存储。RADOS保证CEPH中的各种数据都是对象形式，操持CEPH的一致性。

RBD：为客户端提供块存储接口

RADOS GW：为客户端提供对象存储接口

CEPH FS：为客户端提供文件系统存储接口

CEPH环境搭建

1. 创建5台虚拟机

node1.tedu.cn 192.168.4.1

node2.tedu.cn 192.168.4.2

node3.tedu.cn 192.168.4.3

node4.tedu.cn 192.168.4.4

client.tedu.cn 192.168.4.10

2、启动虚机

[root@room8pc16 kvms\_ansi]# for vm in rh7\_node{1..5}

> do

> virsh start $vm

> done

1. 在物理主机上配置CEPH的YUM源

[root@room8pc16 cluster]# mkdir /var/ftp/ceph/

[root@room8pc16 cluster]# tail -1 /etc/fstab

/ISO/rhcs2.0-rhosp9-20161113-x86\_64.iso /var/ftp/ceph iso9660 defaults 0 0

[root@room8pc16 cluster]# mount -a

[root@room8pc16 ~]# vim 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

1. CEPH集群中有很多节点，逐台管理效率低下，还可能出现错误，所以可以找一台主机当作管理节点，由它统一管理所有主机。我们保用node1作为管理节点。
2. 为了使得管理节点能够方便的管理，先创建免密登陆
3. 通过名称访问各台主机，配置名称解析

[root@node1 ~]# for i in {1..4}

> do

> echo -e "192.168.4.$i\tnode$i.tedu.cn\tnode$i" >> /etc/hosts

> done

[root@node1 ~]# echo -e "192.168.4.10\tclient.tedu.cn\tclient" >> /etc/hosts

1. 生成密钥对，非交互模式生成

[root@node1 ~]# ssh-keygen -f /root/.ssh/id\_rsa -N ""

1. 第一交ssh到远程主机会被询问(yes/no)?。可以先将远程主机的身份信息，保存到本地

[root@node1 ~]# ssh-keyscan 192.168.4.{1..4} >> /root/.ssh/known\_hosts

[root@node1 ~]# ssh-keyscan 192.168.4.10 >> /root/.ssh/known\_hosts

[root@node1 ~]# ssh-keyscan node{1..4} >> /root/.ssh/known\_hosts

[root@node1 ~]# ssh-keyscan client >> /root/.ssh/known\_hosts

（4）拷贝密钥到远程主机

[root@node1 ~]# for ip in 192.168.4.{1..4}

> do

> ssh-copy-id -i $ip

> done

[root@node1 ~]# ssh-copy-id -i 192.168.4.10

（5）把hosts文件拷贝到各台主机

[root@node1 ~]# for host in node{2..4}

> do

> scp /etc/hosts $host:/etc/

> done

[root@node1 ~]# scp /etc/hosts client:/etc

1. 将client作为NTP服务器

NTP：网络时间协议，udp123端口。用于同步时间。

精确时间的确定：原子钟。全球时间不是一样的，因为地球是圆的，所以将地球按经度，每隔15度角划分一个时区，一共24时区。中国采用东八区时间。

1. 在client上安装软件包

[root@client ~]# yum install -y chrony

1. 修改配置

[root@client ~]# vim /etc/chrony.conf

allow 192.168.4.0/24

local stratum 10

1. 启动服务

[root@client ~]# systemctl restart chronyd; systemctl enable chronyd

1. 将其他主机作为客户端

[root@node1 ~]# vim /etc/chrony.conf

server 192.168.4.10 iburst #其他3行server开头的删除

[root@node1 ~]# for ip in 192.168.4.{2..4}

> do

> scp /etc/chrony.conf $ip:/etc/

> done

[root@node1 ~]# for ip in 192.168.4.{1..4}

> do

> ssh $ip systemctl restart chronyd

> done

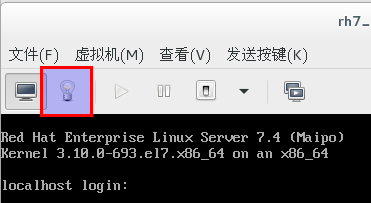
1. 测试

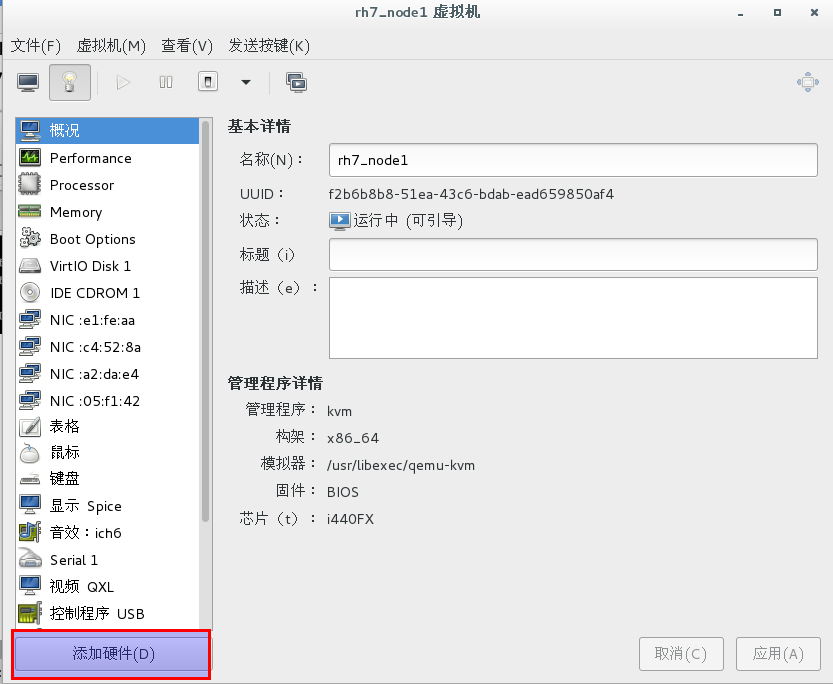
[root@node1 ~]# date -s "2018-06-20 12:00:00"

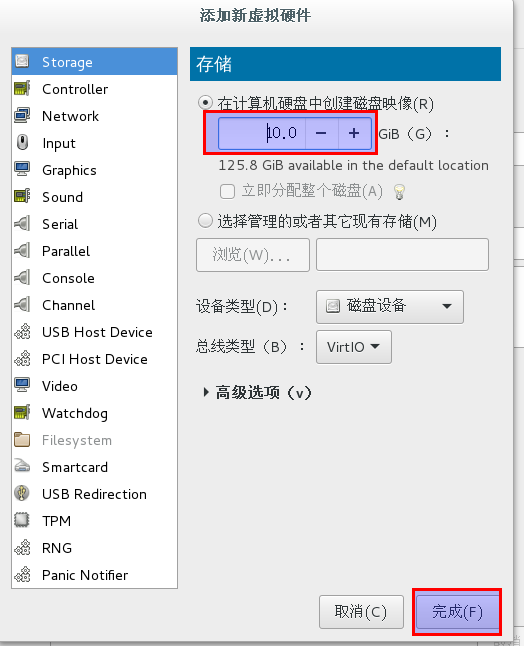
[root@node1 ~]# ntpdate 192.168.4.10 跟192.168.4.10同步时钟

[root@node1 ~]# date 时间已同步

1. 在node1~node3上各添加3块硬盘







1. 在node1节点上安装ceph部署工具

[root@node1 ~]# yum install -y ceph-deploy

1. 在node1节点上为ceph创建工作目录，目录名自定义

[root@node1 ~]# mkdir ceph\_conf

[root@node1 ~]# cd ceph\_conf

1. 生成ceph安装所必须的配置文件

[root@node1 ceph\_conf]# ceph-deploy new node1 node2 node3

[root@node1 ceph\_conf]# ls

1. 安装ceph集群

[root@node1 ceph\_conf]# ceph-deploy install node1 node2 node3

1. 初始化所有节点的MON服务

[root@node1 ceph\_conf]# ceph-deploy mon create-initial

配置CEPH集群

1. 在node1~node3节点上对vdb进行分区，将vdb的分区用于日志

[root@node1 ceph\_conf]# for host in node{1..3}

> do

> ssh $host parted /dev/vdb mklabel gpt

> done

[root@node1 ceph\_conf]# for host in node{1..3}; do ssh $host parted /dev/vdb mkpart primary 1024kB 50%; done

[root@node1 ceph\_conf]# for host in node{1..3}; do ssh $host parted /dev/vdb mkpart primary 50% 100%; done

[root@node1 ceph\_conf]# for host in node{1..3}; do ssh $host lsblk; done

[root@node1 ceph\_conf]# for host in node{1..3}; do ssh $host chown ceph.ceph /dev/vdb?; done # 系统重启后属主属组又变回root.disk了。

配置udev，使得磁盘属主属组在reboot后，仍然是ceph

[root@node3 ~]# vim /etc/udev/rules.d/90-mydisk.rules

ACTION=="add", KERNEL=="vdb[12]", OWNER="ceph", GROUP="ceph"

1. 创建OSD磁盘，在node1上执行
2. 初始化磁盘

[root@node1 ceph\_conf]# for host in node{1..3}

> do

> ceph-deploy disk zap $host:vdc $host:vdd

> done

1. 创建OSD，将数据的日志区指定为vdb

[root@node1 ceph\_conf]# for host in node{1..3}

> do

> ceph-deploy osd create $host:vdc:/dev/vdb1 $host:vdd:/dev/vdb2

> done

如果出现run ‘gatherkeys’的错误提示，执行以下命令

[root@node1 ceph\_conf]# ceph-deploy gatherkeys node1 node2 node3

1. 查看状态

[root@node1 ceph\_conf]# ceph -s 如果正常将显示HEATH\_OK

如果状态是HEALTH\_ERR，就重启服务，如下：

[root@node1 ceph\_conf]# for host in node{1..3}; do ssh $host systemctl restart ceph\\*.service ceph\\*.target ; done