**存储服务器集群的综合练习**

准备：

client50：192.168.4.50 做客户机，做成NTP服务器 ,其他主机以50为NTP

// echo “allow 192.168.4.0/24’ > /etc/chrony.conf

node51：192.168.4.51 加三块10G的硬盘

node52：192.168.4.52 加三块10G的硬盘

node53：192.168.4.53 加三块10G的硬盘

node54：192.168.4.54

搭建源：真机共享mount /iso/rhcs2.0-rhosp9-20161113-x86\_64.iso /var/ftp/ceph

/var/ftp/ceph/rhceph-2.0-rhel-7-x86\_64/MON/

/var/ftp/ceph/rhceph-2.0-rhel-7-x86\_64/OSD/

/var/ftp/ceph/rhceph-2.0-rhel-7-x86\_64/Tools/

cat /etc/hosts //书写hosts文件，每台主机都要有这个配置

192.168.4.50 client50

192.168.4.51 node51

192.168.4.52 node52

192.168.4.53 node53

node51：无密码连接 client50、node51、node52、node53

ssh-keygen -f /root/.ssh/id\_rsa -N ‘’ //非交互式生成密钥对

将公钥发给其他主机及自己，实现ssh无密码登陆

for i in 51 52 53; do ssh-copy-id 192.168.4.$i ; done

**分布式文件系统**

分布式文件系统（Distributed File System）是指文件系统管理的物理存储资源不一定直接连接再本地节点上，而是通过计算机网络与节点相连

分布式文件系统的设计基于客户机/服务器（C/S）模式

常用分布式文件系统

Lustre，Hadoop，FastDFS，Ceph，GlusterFS

ceph简介

有官方(付费)的和开源的

Ceph 是一个分布式文件系统

具有高扩展、高可用、高性能的特点

**Ceph 可以提供**对象存储、块存储、文件系统存储

Ceph 可以提供PB级别的存储空间（PB-->TB-->-->GB）

软件定义存储（Software Defined Storage）作为存储，行业的一大发展趋势

官网：<http://docs.ceph.org/start/intro>

ceph 组件

OSDs ：存储设备

Monitors ：集群监控组件

MDSs ：存放文件系统的元数据（对象存储和块存储不需要该组件）

元数据：文件的信息，大小，权限等，即如下信息

drwxr-xr-x 2 root root 6 10月 11 10:37 /root/a.sh

Client ：ceph客户端

实验：

使用node51作为**部署主机，管理主机**

node51：

1.安装部署软件ceph-deploy：

yum -y install ceph-deploy //安装完成后使用ceph-deploy --help帮助

为部署工具创建目录，存放密钥与配置文件

//安装完成后会

mkdir /root/ceph-cluster

**cd /root/ceph-cluster //以后的操作尽量都再这个文件夹下进行**

2.创建Ceph集群

创建Ceph集群配置（所有节点都为mon）

ceph-deploy new node51 node52 node53

给所有节点安装Ceph软件包

ceph-deploy install node51 node52 node53

初始化所有节点的mon(监控程序)服务 （每一台主机上都配置有/etchosts主机名解析）

ceph-deploy mon create-initial

3.创建OSD

I所有节点准备磁盘分区（以node51为例，node52，node53也要这样分区）

Node51 以下的分区为日志磁盘分区

1)指定用那种分区模式

parted /dev/vdb mklabel gpt

2)用这个盘的前50%的空间建一个分区，以1M为起点

parted /dev/vdb mkpart primary 1M 50%

3)用这个盘的后50%的空间建一个分区

parted /dev/vdb mkpart primary 50% 100%

扩展：parted /dev/vdb rm 1 //删除分区，分区号（id）

4)将这两个分区的所有者和所有组设置为ceph，给集群ceph管理权限

chown ceph.ceph /dev/vdb1

chown ceph.ceph /dev/vdb1

echo 'chown ceph.ceph /dev/vdb\*' >> /etc/rc.d/rc.local

chmod +x /etc/rc.d/rc.local

注：这两个分区用来做存储服务器的日志journal磁盘分区

II初始化清空磁盘数据（仅在node51上进行管理操作）

cd /root/ceph-cluster/ //必须再这个目录下进行操作

ceph-deploy disk zap node51:vdc node51:vdd

ceph-deploy disk zap node52:vdc node52:vdd

ceph-deploy disk zap node53:vdc node53:vdd

III创建OSD存储设备（仅在node51上进行管理操作）

ceph-deploy osd create node51:vdc:/dev/vdb1

node51:vdd:/dev/vdb2 //同时指定日志journal磁盘

>>Host node51 is now ready for osd use.

ceph-deploy osd create node52:vdc:/dev/vdb1

node52:vdd:/dev/vdb2

>>Host node52 is now ready for osd use.

ceph-deploy osd create node53:vdc:/dev/vdb1

node53:vdd:/dev/vdb2

>>Host node53 is now ready for osd use.

IIII 重启所有的ceph服务

systemctl restart ceph\\*.service ceph\\*.target

//node51、node52、node53这些服务应该都是正常运行的，既做存储也做管理，\*.target管理服务,\*.servertigongfuwu

**服务查看**

**Node51: 有关与ceph的服务**

ceph-create-keys@node51.service ceph-osd@1.service

ceph-mds.target ceph-osd.target

ceph-mon@node51.service ceph-radosgw.target

ceph-mon.target ceph.target

ceph-osd@0.service

**Node52: 有关与ceph的服务**

ceph-create-keys@node52.service ceph-osd@2.service

ceph-disk@dev-vdd1.service ceph-osd.target

ceph-mds.target ceph-radosgw.target

ceph-mon@node52.service ceph.target

ceph-mon.target

**Node53: 有关与ceph的服务**

ceph-create-keys@node53.service ceph-osd@4.service

ceph-disk@dev-vdd1.service ceph-osd.target

ceph-mds.target ceph-radosgw.target

ceph-mon@node53.service ceph.target

ceph-mon.target

**部署ceph集群**

安装部署软件ceph-deploy

创建Ceph集群

创建OSD存储空间

查看ceph状态，验证

**块存储**

单机块设备：光盘、磁盘

分布式块存储：ceph、cider

Ceph块设备也叫RADOS块设备

- RADOS block device：RBD

Rbd驱动已经很好的集成在了linux内核中

Rbd提供了企业功能。如快照、COW（Copy Online Write,写时复制）克隆

COW对源文件做写操作时，旧数据会被复制到快照文件里。当删除文件或者对文件进行了内容的增加或者减少，源文件发生了改变，旧的文件就会拷贝到快照里

Rbd还支持内存缓存，从而能够大大提高性能

**块存储集群(分配指定大小的镜像)**

镜像池大小，基本存储为60G，为node51、node52、node53做存储磁盘的和

查看存储池（默认有一个rbd池）

ceph osd lspools

0 rbd

创建镜像, ##若不指定存储池，默认属于rbd存储池

rbd create demo-image --image-feature layering --size 10G

//镜像名为demo-image , --image-feature layering(创建镜像的方式)

默认在rbd默认的存储池里创建

rbd create rbd/image --image-feature layering --size 10G

//rbd/image 指定在rbd池里创建

查看镜像

rbd info demo-image

rbd image 'demo-image':

size 10240 MB in 2560 objects

order 22 (4096 kB objects)

block\_name\_prefix: rbd\_data.1052238e1f29

format: 2

features: layering

flags:

删除镜像，

rbd remove rbd/image //从指定的池中删除镜像

rbd remove demo-image //次出仅提供错误时的方法，不实际执行

删除时需要确认其没有快照，若有快照，使用以下命令清除快照

rbd snap purge image

再执行 rbd remove rbd/image

缩小容量 //命令解释，重置大小为1G 镜像image 允许缩小

rbd resize --size 1G image --allow-shrink

扩大容量 //将容量扩大到2G

rbd resize --size 2G image

**集群内**通过rbd访问

1.node51本机使用，将镜像映射为本地磁盘

rbd map demo-image

lsblk //查看本地磁盘

rbd0 251:0 0 10G 0 disk

2.分区、格式化（分区名为/dev/rbd0p1）、挂载，与本地磁盘无异

3.将镜像从本地磁盘中移除 //移除前需要将挂载从本机卸载掉

rbd unmap demo-image

**集群外**客户机clinet50：通过rbd访问

1.安装ceph-common软件包

yum -y install ceph-common.x86\_64

2.拷贝配置文件（指示储存集群的位置）

scp 192.168.4.51:/etc/ceph/ceph.conf /etc/ceph/

3.拷贝连接接密钥（获取集群连接及使用的权限）

scp 192.168.4.51:/etc/ceph/ceph.client.admin.keyring /etc/ceph/

1. 查看集群镜像

Rbd list

Demo-image

Image

1. 镜像映射到本地磁盘

rbd map image

lsblk

6.显示本地影射

rbd showmapped

id pool image snap device

0 rbd image - /dev/rbd0

7.分区、格式化（分区名为/dev/rbd0p1）、挂载，与本地磁盘无异

8.撤销磁盘映射,将镜像从本地磁盘中移除 //移除前需要将挂载从本机卸载掉

rbd unmap demo-image

rbd unmap /dev/rbd/rbd/demo-image //两种方法

**创建镜像快照** 16:20

快照使用COW技术，对大数据快照速度会很快

COW对源文件做写操作时，旧数据会被复制到快照文件里

快照：保存某一时刻的所有信息，以备以后恢复使用，创建初期不占用磁盘空间，每当 源文件发生改变时，就把快照创建时的文件数据写入快照，这时快照开始占用磁 盘空间，大小为发生改变的文件大小的和。

node51:

查看已有镜像

rbd list

查看镜像快照：

rbd snap ls image //暂时无显示

**创建快照（snap）**

rbd snap create image --snap image-snap1

命令解释：rbd 快照 create 镜像名 --snap类型 快照名

再次查看镜像快照

rbd snap ls image

SNAPID NAME SIZE

4 image-snap1 2048 MB

使用快照恢复数据

rbd snap rollback image --snap image-snap1

客户机将镜像卸载后再重新挂载即可恢复数据

删除快照

Rbd snap rm image --snap image-snap1

**快照克隆**

- 如果想从快照恢复出来一个新的镜像，可以做克隆

- 注意，克隆前，需要对快照进行<保护>操作

- 被保护的快照无法删除，取消保护

快照保护：

rbd snap protect image --snap image-snap1

快照克隆

rbd clone image --snap image-snap1 image-clone --image-feature layering

克隆查看

rbd info image-clone

rbd image 'image-clone':

size 2048 MB in 512 objects

order 22 (4096 kB objects)

block\_name\_prefix: rbd\_data.108a2ae8944a

format: 2

features: layering

flags:

parent: rbd/image@image-snap1 //父从属信息

overlap: 2048 MB

利用克隆镜像恢复数据,

rbd flatten image-clone

再次查看克隆镜像

rbd info image-clone

rbd image 'image-clone':

size 2048 MB in 512 objects

order 22 (4096 kB objects)

block\_name\_prefix: rbd\_data.108a2ae8944a

format: 2

features: layering

flags:

取消保护：//必须取消保护才能撤销磁盘映射

rbd snap unprotect image --snap image-snap1

客户端撤销磁盘映射

1.卸载挂载点

2.查看rbd磁盘映射

Rbd showmapped

id pool image snap device

0 rbd image - /dev/rbd0

3.撤销磁盘映射

rbd unmap demo-image

rbd unmap /dev/rbd/rbd/demo-image //两种方法

再听一遍，11：10

Fdisk分区单个文件不能超过2G

Exit4日志文件系统

这种分区一部分用来存储日志，一部分存储数据，有几个数据存储的数据盘，就必须要有几个存储日志的日志盘