

CLUSTER DAY04



集群与存储

NSD CLUSTER

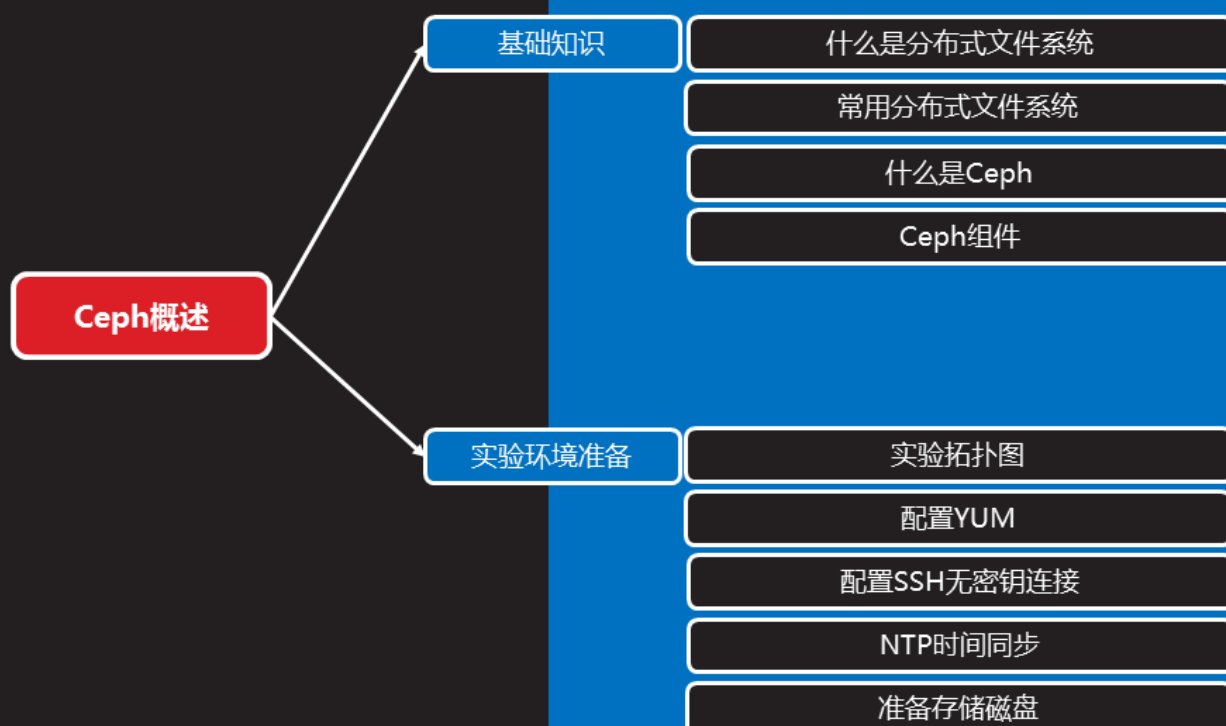
DAY04

内容

上午	09:00 ~ 09:30	作业讲解和回顾
	09:30 ~ 10:20	Ceph概述
	10:30 ~ 11:20	
	11:30 ~ 12:20	部署Ceph集群
下午	14:00 ~ 14:50	
	15:00 ~ 15:50	Ceph块存储
	16:10 ~ 17:00	
	17:10 ~ 18:00	总结和答疑



Ceph概述



基础知识



什么是分布式文件系统

- 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在本地节点上，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/服务器模式

常用分布式文件系统

知识讲解

- Lustre
- Hadoop
- FastDFS
- Ceph
- GlusterFS



什么是Ceph

知识讲解

- 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)
 - $1024G \times 1024G = 1048576G$
- 软件定义存储(Software Defined Storage)作为存储行业的一大发展趋势，已经越来越受到市场的认可

帮助文档：<http://docs.ceph.org/start/intro>



Ceph组件

知识讲解

- OSDs
 - 存储设备
- Monitors
 - 集群监控组件
- RBD
 - 对象存储网关
- 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）
- Client
 - ceph客户端

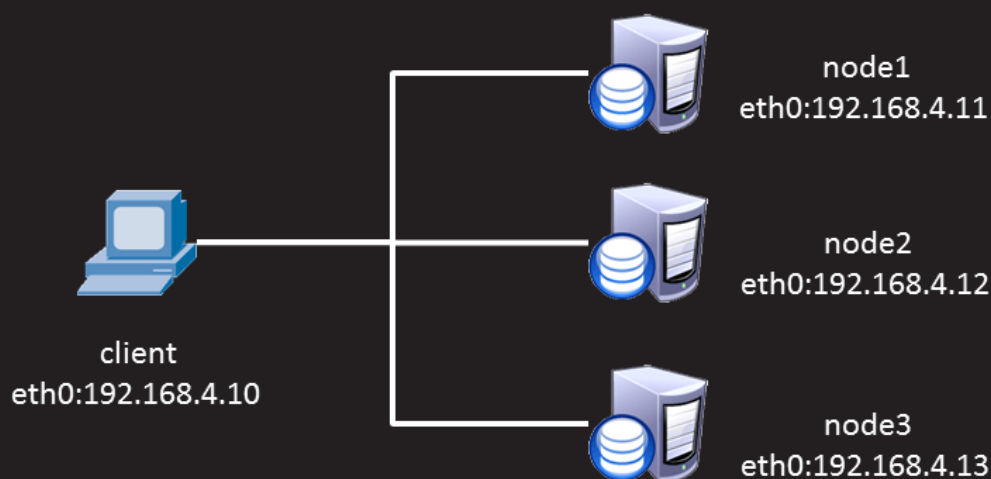


实验环境准备

实验拓扑图

- 1台客户端虚拟机
- 3台存储集群虚拟机

知识讲解



配置YUM

- 物理机创建网络yum源服务器

```
[root@root9pc01 ~]# yum -y install vsftpd
[root@root9pc01 ~]# mkdir /var/ftp/ceph
[root@root9pc01 ~]# mount -o loop \
rhcs2.0-rhosp9-20161113-x86_64.iso /var/ftp/ceph
[root@root9pc01 ~]# systemctl restart vsftpd
```

知识讲解



配置YUM (续1)

- 虚拟机调用YUM源 (下面以node1为例)

```
[root@node1 ~]# cat /etc/yum.repos.d/ceph.repo
[mon]
name=mon
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/MON
gpgcheck=0
[osd]
name=osd
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/OSD
gpgcheck=0
[tools]
name=tools
baseurl=ftp://192.168.4.254/ceph/rhceph-2.0-rhel-7-x86_64/Tools
gpgcheck=0
```

知识讲解



配置SSH无密钥连接

- 修改主机名
- 警告：/etc/hosts解析的域名必须与本机主机名一致！！

```
[root@node1 ~]# cat /etc/hosts
... ..
192.168.4.10    client
192.168.4.11    node1
192.168.4.12    node2
192.168.4.13    node3
[root@node1 ~]# for i in 10 11 12 13
> do
> scp /etc/hosts 192.168.2.$i:/etc/
> done
```

知识讲解



配置SSH无密钥连接（续1）

- 非交互生成密钥对

```
[root@node1 ~]# ssh-keygen -f /root/.ssh/id_rsa -N ""
```

- 发布密钥到各个主机（包括自己）

```
[root@node1 ~]# for i in 10 11 12 13
> do
> ssh-copy-id 192.168.4.$i
> done
```

知识讲解



NTP时间同步

- 客户端创建NTP服务器

```
[root@client ~]# yum -y install chrony
[root@client ~]# cat /etc/chrony.conf
server 0.centos.pool.ntp.org iburst
allow 192.168.4.0/24
local stratum 10
[root@client ~]# systemctl restart chronyd
```

- 其他所有主机与其同步时间（下面以node1为例）

```
[root@node1 ~]# cat /etc/chrony.conf
server 192.168.4.10 iburst
[root@node1 ~]# systemctl restart chronyd
```

知识讲解



准备存储磁盘

知识讲解

- 物理机上为每个虚拟机创建3个磁盘

```
[root@root9pc01 ~]# cd /var/lib/libvirt/images
[root@root9pc01 ~]# qemu-img create -f qcow2 node1-vdb.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node1-vdc.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node1-vdd.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node2-vdb.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node2-vdc.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node2-vdd.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node3-vdb.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node3-vdc.vol 10G
[root@root9pc01 ~]# qemu-img create -f qcow2 node3-vdd.vol 10G
```
- 在图形环境中为虚拟机添加磁盘

```
[root@root9pc01 ~]# virt-manager
```

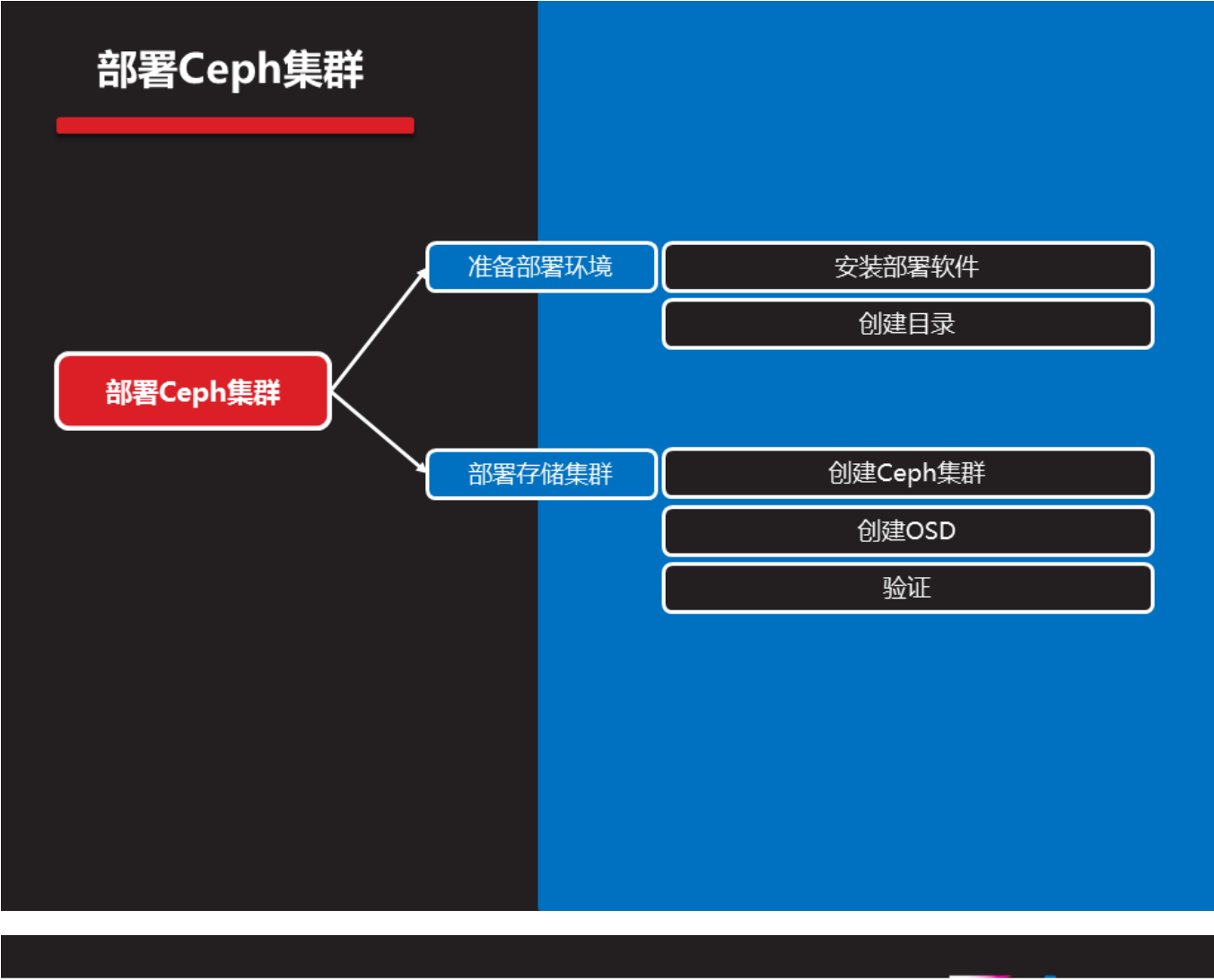


案例1：实验环境

课堂练习

- 创建1台客户端虚拟机
- 创建3台存储集群虚拟机
- 配置主机名、IP地址、YUM源
- 修改所有主机的主机名
- 配置无密码SSH连接
- 配置NTP时间同步
- 创建虚拟机磁盘





安装部署软件

- 使用node1作为部署主机

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

知识讲解

- ceph-deploy命令与子命令都支持--help查看帮助

```
[root@node1 ~]# ceph-deploy --help
```



创建目录

- 为部署工具创建目录，存放密钥与配置文件

```
[root@node1 ~]# mkdir ceph-cluster
```

```
[root@node1 ~]# cd ceph-cluster/
```

知识讲解



部署存储集群

创建Ceph集群

知识讲解

- 创建Ceph集群配置（所有节点都为mon）
`[root@node1 ceph-cluster]# ceph-deploy new node1 node2 node3`
- 给所有节点安装Ceph软件包
`[root@node1 ceph-cluster]# ceph-deploy install node1 node2 node3`
- 初始化所有节点的mon服务（主机名解析必须对）
`[root@node1 ceph-cluster]# ceph-deploy mon create-initial`
//这里没有指定主机，是因为第一步创建的配置文件中已经有了，
//所以要求主机名解析必须对，否则连接不到对应的主机



创建Ceph集群（续1）

- 初始化monitor常见错误

如果提示如下错误信息：

```
[node1][ERROR] admin_socket: exception getting command descriptions:  
[Error 2] No such file or directory
```

解决方案如下（在node1操作）：

```
[root@node1 ceph-cluster]# vim ceph.conf #文件最后追加以下内容  
public_network = 192.168.4.0/24
```

修改后重新推送配置文件

```
[root@node1 ceph-cluster]# ceph-deploy --overwrite-conf config push  
node1 node2 node3
```

知识
讲解



创建OSD (续1)

知识讲解

- 初始化清空磁盘数据 (仅node1操作即可)

```
[root@node1 ~]# ceph-deploy disk zap node1:vdc node1:vdd
```

```
[root@node1 ~]# ceph-deploy disk zap node2:vdc node2:vdd
```

```
[root@node1 ~]# ceph-deploy disk zap node3:vdc node3:vdd
```

- 创建OSD存储空间 (仅node1操作即可)

```
[root@node1 ~]# ceph-deploy osd create node1:vdc:/dev/vdb1  
node1:vdd:/dev/vdb2
```

//创建osd存储设备，vdc为集群提供存储空间，vdb1提供JOURNAL日志，一个存储设备对应一个日志设备，日志需要SSD，不需要很大

```
[root@node1 ~]# ceph-deploy osd create node2:vdc:/dev/vdb1  
node2:vdd:/dev/vdb2
```

```
[root@node1 ~]# ceph-deploy osd create node3:vdc:/dev/vdb1  
node3:vdd:/dev/vdb2
```



验证（续2）

知识讲解

- 查看集群状态

```
[root@node1 ~]# ceph -s
```

- 可能出现的错误

- health: HEALTH_WARN

clock skew detected on node2, node3...

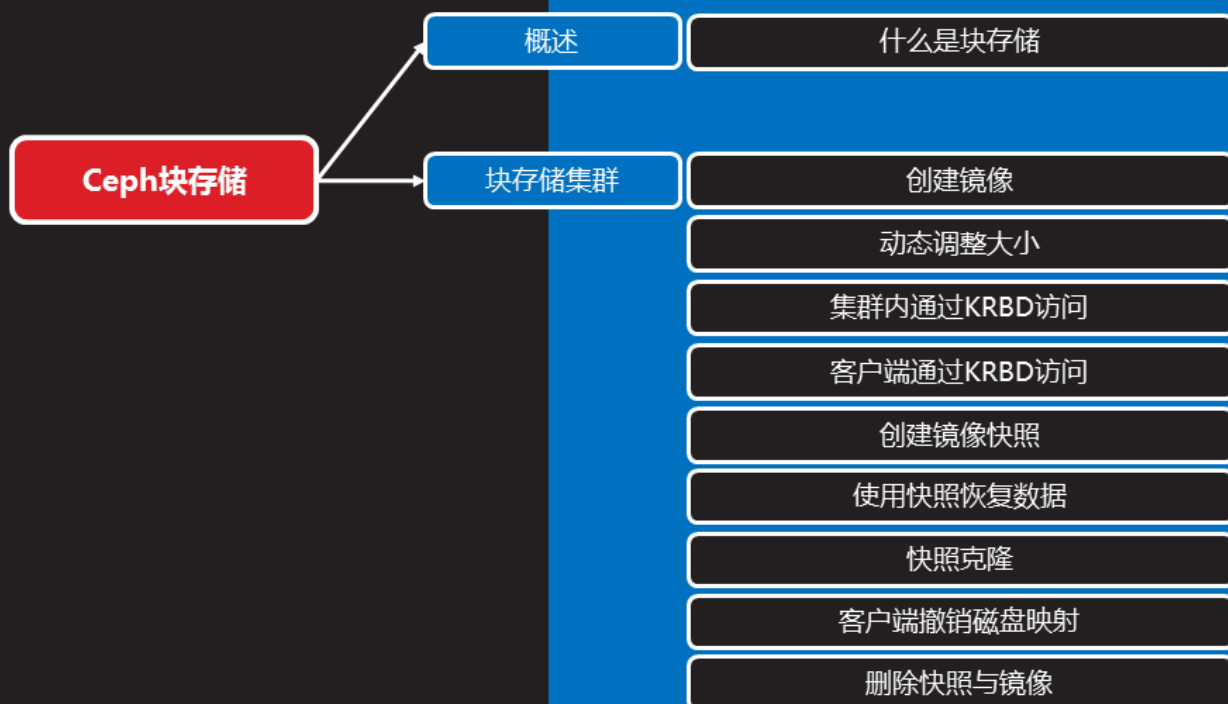
- clock skew(时间不同步)

- 解决：

- 请先将所有主机的时间都使用NTP时间同步！！！！



Ceph块存储



概述

