Tedu.cn
达内教育

集群与存储

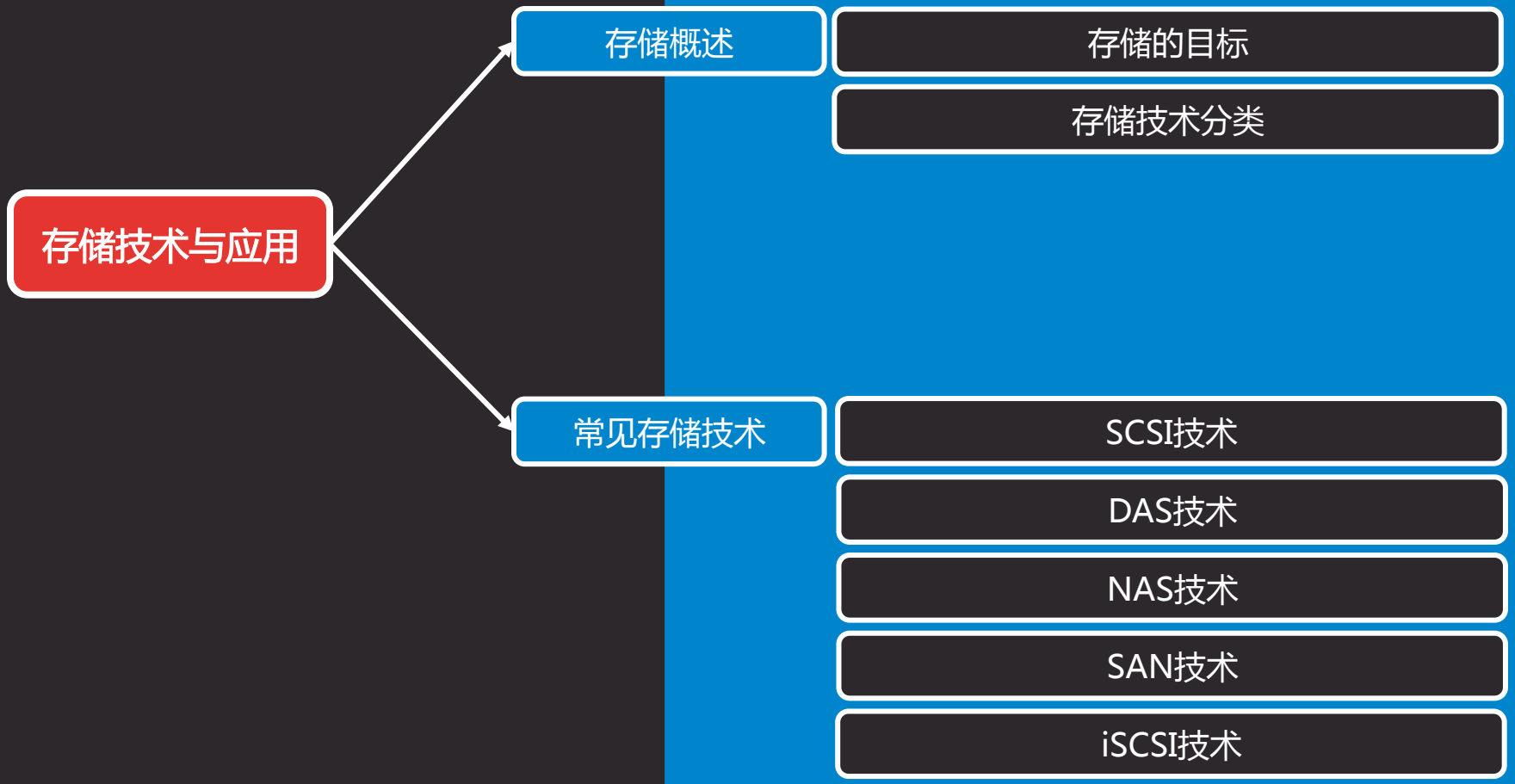
NSD CLUSTER

DAY01

内容

上午	09:00 ~ 09:30	存储技术与应用
	09:30 ~ 10:20	iSCSI技术应用
	10:30 ~ 11:20	
	11:30 ~ 12:20	udev配置
下午	14:00 ~ 14:50	
	15:00 ~ 15:50	NFS网络文件系统
	16:10 ~ 17:00	Multipath多路径
	17:10 ~ 18:00	总结和答疑

存储技术与应用


存储技术与应用
存储概述
存储的目标
存储技术分类
常见存储技术
SCSI技术
DAS技术
NAS技术
SAN技术
iSCSI技术

Tedu.cn
达内教育

存储概述

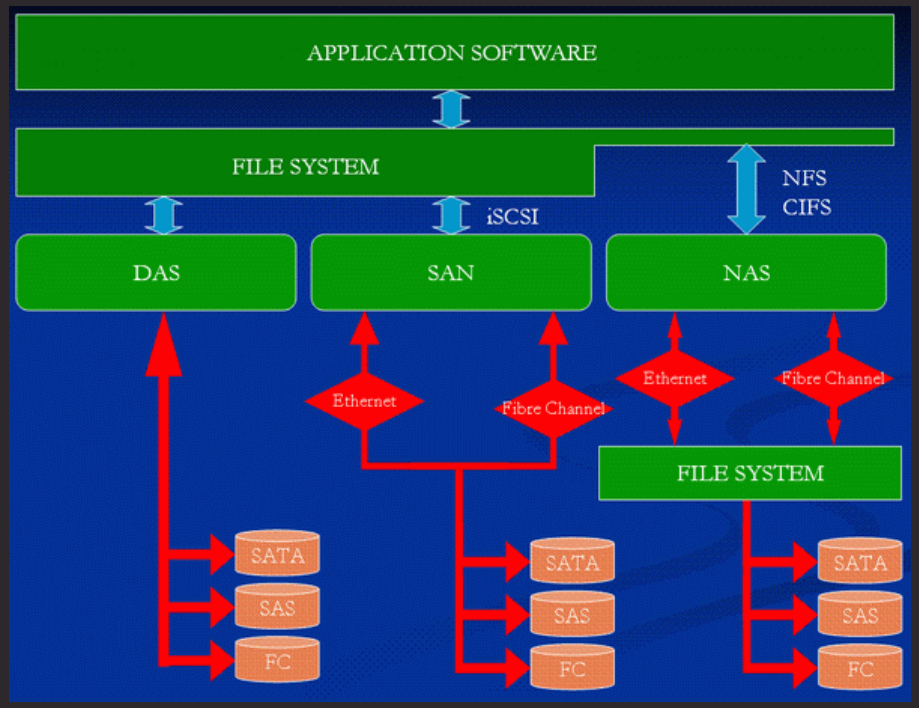
存储的目标

- 存储是根据不同的应用环境通过采取合理、安全、有效的方式将数据保存到某些介质上并能保证有效的访问
- 一方面它是数据临时或长期驻留的物理媒介
- 另一方面，它是保证数据完整安全存放的方式或行为
- 存储就是把这两个方面结合起来，向客户提供一套数据存放解决方案

知识讲解

存储技术分类

- SCSI小型计算机系统接口
- DAS直连式存储
- NAS网络技术存储
- SAN存储区域网络
- FC光纤通道

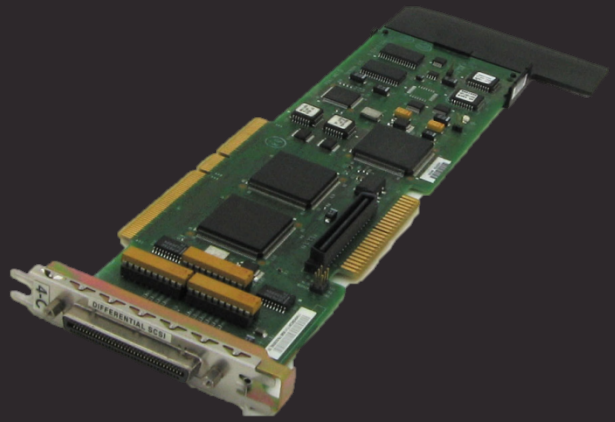


Tedu.cn
达内教育

常见存储技术

SCSI技术

- Small Computer System Interface的简称
- 作为输入/输出接口
- 主要用于硬盘、光盘、磁带机等设备



知识讲解

DAS技术

- Direct-Attached Storage的简称
- 将存储设备通过SCSI接口或光纤通道直接连接到计算机上
- 不能实现数据与其他主机的共享
- 占用服务器操作系统资源，如CPU、IO等
- 数据量越大，性能越差

知识讲解

NAS技术

- Network-Attached Storage的简称
- 一种专用数据存储服务器，以数据为中心，将存储设备与服务器彻底分离，集中管理数据，从而释放带宽、提高性能、降低总拥有成本、保护投资
- 用户通过TCP/IP协议访问数据
  - 采用标准的NFS/HTTTP/CIFS等

知识讲解

SAN技术

- Storage Area Network的简称
  - 通过光纤交换机、光纤路由器、光纤集线器等设备将磁盘阵列、磁带等存储设备与相关服务器连接起来，形成高速专网网络
- 组成部分
  - 如路由器、光纤交换机
  - 接口：如SCSI、FC
  - 通信协议：如IP、SCSI

知识讲解

SAN技术（续1）

- Fibre Channel
  - 一种适合于千兆数据传输的、成熟而安全解决方案
  - 与传统的SCSI相比，FC提供更高的数据传输速率、更远的传输距离、更多的设备连接支持以及更稳定的性能、更简易的安装

知识讲解

SAN技术（续2）

- FC主要组件
  - 光纤
  - HBA（主机总线适配置器）
  - FC交换机

知识讲解

SAN技术（续3）

- FC交换机交换拓扑
  - 点到点：point-to-point
    简单将两个设备互连
  - 已裁定的环路：arbitrated loop
    可多达126个设备共享一段信道或环路
  - 交换式拓扑：switched fabric
    所有设备通过光纤交换机互连

知识讲解

iSCSI技术

- Internet SCSI
- IETF制定的标准，将SCSI数据块映射为以太网数据包
- 是一种基于IP Storage理论的新型存储技术
- 将存储行业广泛应用的SCSI接口技术与IP网络相结合
- 可以在IP网络上构建SAN
- 最初由Cisco和IBM开发

知识讲解

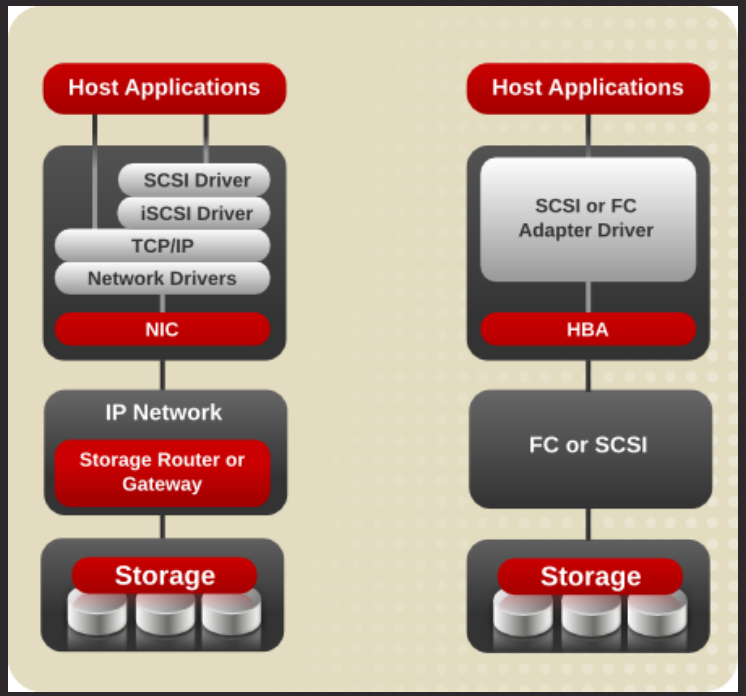
iSCSI技术（续1）

- 优势
  - 基于IP协议技术的标准
  - 允许网络在TCP/IP协议上传输SCSI命令
  - 相对FC SAN，iSCSI实现的IP SAN投资更低
  - 解决了传输效率、存储容量、兼容性、开放性、安全性等方面的问题
  - 没有距离限制

知识讲解

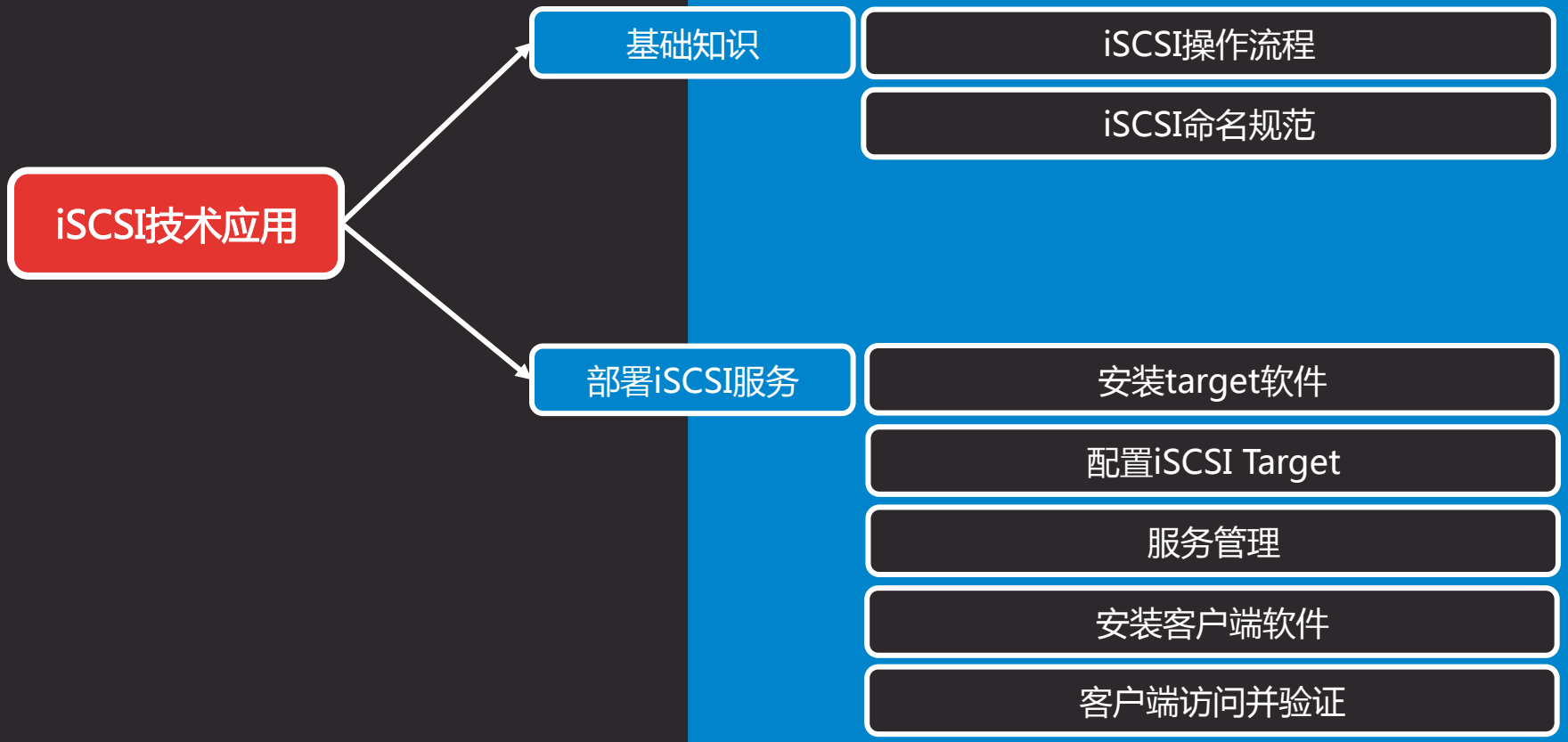
iSCSI技术（续2）

- 客户端
  - iSCSI Initiator：软件实现，成本低、性能较低
  - iSCSI HBA：硬件实现，性能好，成本较高
- 存储设备端
  - iSCSI Target
- 以太网交换机



知识讲解

iSCSI技术应用


iSCSI技术应用
基础知识
iSCSI操作流程
iSCSI命名规范
部署iSCSI服务
安装target软件
配置iSCSI Target
服务管理
安装客户端软件
客户端访问并验证

Tedu.cn
达内教育

基础知识

iSCSI操作流程

- Target端
  - 选择target名称
  - 安装iSCSI target
  - 准备用于target的存储
  - 配置target
  - 启用服务
- Initiator端
  - 安装initiator
  - 配置initiator并启动服务

知识讲解

iSCSI命名规范

- 建议采用IQN（iSCSI限定名称）
- 全称必须全局唯一
- IQN格式：
  iqn.<date_code>.<reversed_domain>.<string>[:<substring>]
- 命名示例：
- iqn.2013-01.com.tarena.tech:sata.rack2.disk1

知识讲解

部署iSCSI服务

安装target软件

- 查询yum仓库

  [root@svr1 ~]# yum list | grep target

- 安装

  [root@svr1 ~]# yum -y install targetcli

- 查看iSCSI target信息

  [root@svr1 ~]# yum info targetcli

知识讲解

配置iSCSI Target

知识讲解

- 定义后端存储

```
[root@svr1 ~]# targetcli
/> ls
/> backstores/block create /dev/vdb1
```

- 创建iqn对象

```
/> /iscsi create iqn.2018-01.cn.tedu:server1
```

- 授权客户机访问

```
/> iscsi/iqn.2018-01.cn.tedu:server1/tpg1/acls create iqn.
2018-01.cn.tedu:client1
```

配置iSCSI Target（续1）

知识讲解

- 绑定存储

  /> iscsi/iqn.2018-01.cn.tedu:server1/tpg1/luns create /backstores/block/iscsi_store

- 绑定监听地址

  /> iscsi/iqn.2018-01.cn.tedu:server1/tpg1/portals/ create 0.0.0.0

- 保存配置

  /> saveconfig
  /> exit

服务管理

- 控制服务

  [root@svr1 ~]# systemctl {start|restart|stop|status} target

- 设置服务开机运行

  [root@svr1 ~]# systemctl enable target

- 查看端口

  [root@svr1 ~]# netstat -tlnp | grep :3260

知识讲解

安装客户端软件

- 查询yum仓库

  [root@svr1 ~]# yum list | grep initiator

- 安装

  [root@svr1 ~]# yum -y install iscsi-initiator-utils

- 查看iSCSI target信息

  [root@svr1 ~]# yum info iscsi-initiator-utils

知识讲解

客户端访问并验证

- 启动服务

  [root@svr1 ~]# service iscsi start

- 设置本机的iqn名称

  vim /etc/iscsi/initiatorname.iscsi
  InitiatorName=iqn.2018-01.cn.tedu:client1

- 发现远程target存储

  [root@svr1 ~]# iscsiadm --mode discoverydb --type sendtargets --portal 192.168.4.1 --discover

知识讲解

客户端访问并验证（续1）

- 登陆target

```
[root@svr1 ~]#  systemctl restart iscsi
[root@svr1 ~]#  lsblk
[root@svr1 ~]# systemctl enable iscsi
[root@svr1 ~]# systemctl enable iscsid
```

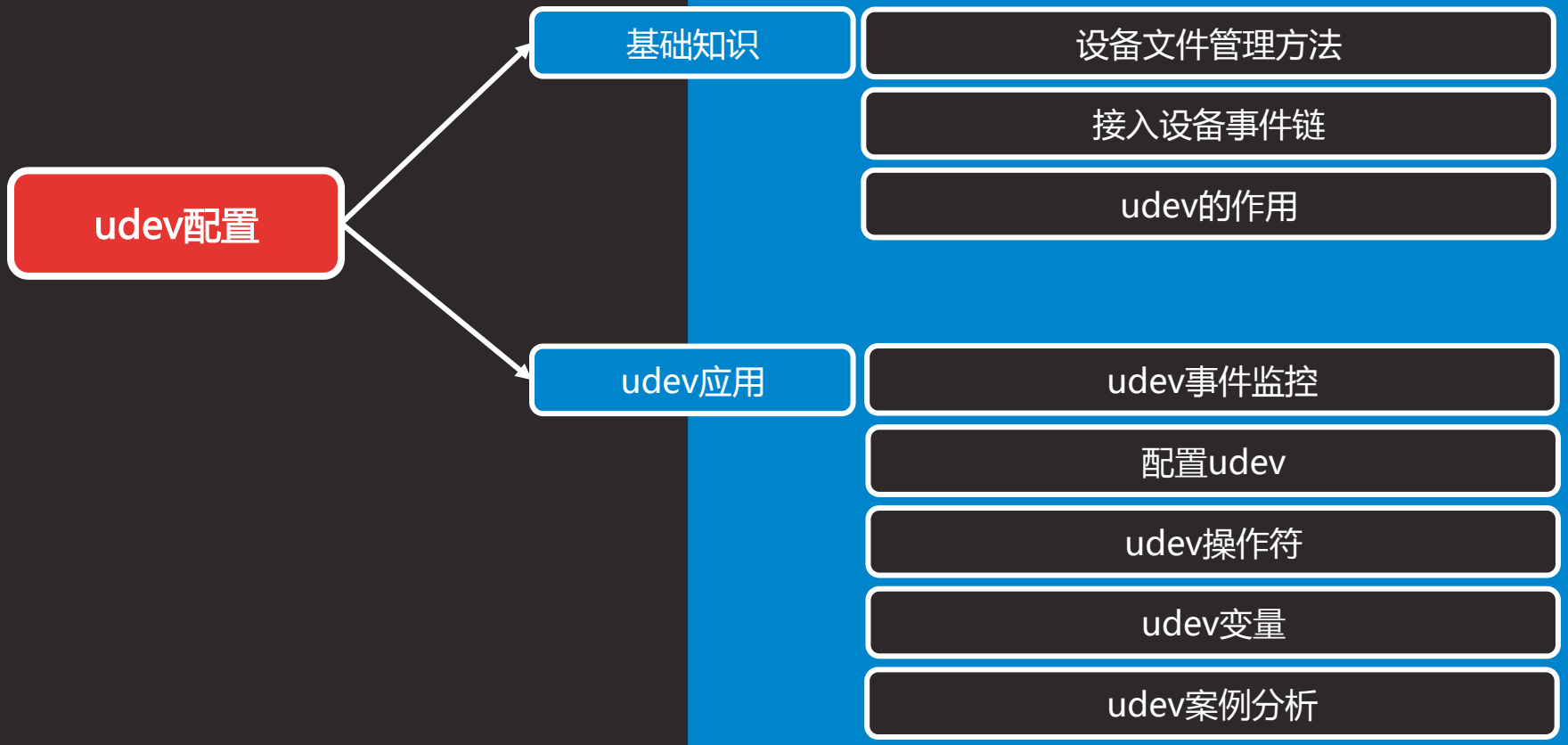
知识讲解

案例1：配置iSCSI服务

- 服务器上要额外配置一块硬盘
- 服务端安装target，并将新加的硬盘配置为iSCSI 的共享磁盘
- 在客户端上安装initiator，挂在服务器iSCSI，要求实现开机自动挂载

课堂练习

udev配置


udev配置
基础知识
设备文件管理方法
接入设备事件链
udev的作用
udev应用
udev事件监控
配置udev
udev操作符
udev变量
udev案例分析

Tedu.cn
达内教育

基础知识

设备文件管理方法

- devfs
  - Linux早期采用的静态管理方法
  - /dev目录下有大量静态文件
  - 内核版本2.6.13开始被完全取代
- udev
  - 只有连到系统上来的设备才在/dev下创建设备文件
  - 与主、次设备编号无关
  - 为设备提供持久、一致的名字

知识讲解

接入设备事件链

- 内核发现设备并导入设备状态到sysfs
- udev接到事件通知
- udev创建设备节点或是运行指定程序
- udev通知hald守护进程
- HAL探测设备信息
- HAL创建设备对象结构
- HAL通过系统消息总线广播该事件
- 用户程序也可以监控该事件

知识讲解

udev的作用

- 从内核收到添加/移除硬件事件时，udev将会分析：
  - /sys目录下信息
  - /etc/udev/rules.d目录中的规则
- 基于分析结果，udev会：
  - 处理设备命名
  - 决定要创建哪些设备文件或链接
  - 决定如何设置属性
  - 决定触发哪些事件

知识讲解

udev应用

udev事件监控

知识讲解

```
[root@node4 ~]# udevadm monitor --property
monitor will print the received events for:
UDEV - the event which udev sends out after rule processing
KERNEL - the kernel uevent

KERNEL[1376971904.652851] add      /module/e1000 (module)
UDEV  [1376971904.653279] add      /module/e1000 (module)
KERNEL[1376971905.065797] add      /devices/
pci0000:00/0000:00:11.0/0000:02:01.0/net/eth0 (net)
.. ..
```

配置udev

- 主配置文件/etc/udev/udev.conf
  - udev_root：创建设备文件位置，默认为/dev
  - udev_rules：udev规则文件位置，默认为/etc/udev/rules.d
  - udev_log：syslog优先级，缺省为err

知识讲解

配置udev（续1）

知识讲解

- 文件位置及格式
  - /etc/udev/rules.d/<rule_name>.rules
  - 例：75-custom.rules
- 规则格式
  - <match-key><op><value>[,...]<assignment-key><op>value[,...]
  - BUS=="usb",SYSFS{serial}=="20043512321411d34721",NAME="udisk"

配置udev（续2）

知识讲解

- 操作符
  - ==：表示匹配
  - != ：表示不匹配
- 匹配示例
  - ACTION=="add"
  - KERNEL=="sd[a-z]1"
  - BUS=="scsi"
  - DRIVER!="ide-cdrom"
  - PROGRAM=="myapp.pl",RESULT=="test"

udev操作符

- 操作符
  - =：指定赋予的值
  - +=：添加新值
  - :=：指定值，且不允许被替换
- 示例
  - NAME="udisk"
  - SYMLINK+="data1"
  - OWNER="student"
  - MODE="0600"

知识讲解

udev变量

- 可以简化或缩写规则
  KERNEL=="sda*",SYMLINK+="iscsi%n"
- 常用替代变量
  - %k：内核所识别出来的设备名，如sdb1
  - %n：设备的内核编号，如sda3中的3
  - %p：设备路径，如/sys/block/sdb/sdb1
  - %%：%符号本身

知识讲解

udev案例分析

知识讲解

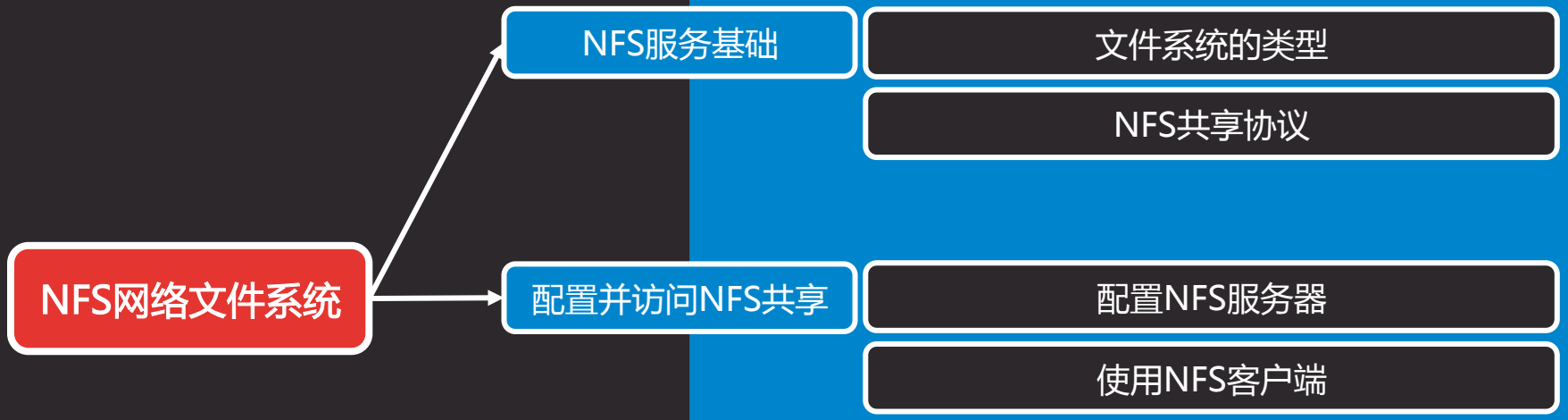
```
SUBSYSTEM=="block",ENV{DEVTYPE}
="disk",KERNEL=="sdb",ENV{ID_VENDOR}
=="TOSHIBA",SYMLINK="udisk",RUN+="/usr/bin/wall udisk plugged
in"
SUBSYSTEM=="block",ACTION=="add",KERNEL=="sdb[0-9]",ENV{ID
_VENDOR_ID}=="0930",ENV{DEVTYPE}=="partition",NAME="udisk
%n"
BUS=="scsi", SYSFS{serial}=="123456789", NAME="byLocation/
rack1-shelf2-disk3"
```

案例2：编写udev规则

- 编写一个插入U盘的规则，要求：
  - 当插入一个U盘时，该U盘自动出现一个链接称为udisk
  - U盘上的第1个分区名称为udisk1，以此类推
  - 终端上出现提示”udisk plugged in”

课堂练习

NFS网络文件系统


NFS网络文件系统
NFS服务基础
文件系统的类型
NFS共享协议
配置并访问NFS共享
配置NFS服务器
使用NFS客户端

Tedu.cn
达内教育

NFS服务基础

文件系统的类型

- 本地文件系统
  - EXT3/4、SWAP、NTFS、……
- 伪文件系统
  - /proc、/sys、……
- 网络文件系统
  - NFS（Network File System）

本地磁盘

内存空间

网络存储空间

知识讲解

NFS共享协议

- Unix/Linux最基本的文件共享机制
  - 1980年由SUN公司开发
  - 依赖于RPC（远程过程调用）映射机制
  - 存取位于远程磁盘中的文档数据，对应用程序是透明的，就好像访问本地的文件一样

知识讲解

Tedu.cn
达内教育

配置并访问NFS共享

配置NFS服务器

- 主要软件包
  - nfs-utils-1.3.0-0.48.el7.x86_64
  - rpcbind-0.2.0-42.el7.x86_64
  - 系统服务脚本
  - nfs、rpcbind
- 主配置文件
  - /etc/exports

知识讲解

配置NFS服务器（续1）

- /etc/exports 配置解析
  - 共享目录 客户机地址(参数, 参数, ...) .. ..

知识讲解

```
[root@svr5 ~]# vim /etc/exports
/root       192.168.4.20(rw)
pc110(rw,no_root_squash)
```

IP地址：192.168.4.20
网段地址：172.0.0.0/24 或 172.0.0.*
所有主机：*
单个域：*.tarena.com
主机名：pc110.tarena.com

rw、ro：可读可写、只读
sync、async：同步写、异步写入
no_root_squash：保留来自客户端的root权限
all_squash：客户端权限都降为nfsnobody

配置NFS服务器（续2）

知识讲解

- 设置2个共享文件夹
  - 将/root 共享给192.168.4.205，可写、保留客户端的root权限
  - 将/usr/src 共享给192.168.4.0/24网段，只读

```
[root@svr5 ~]# vim  /etc/exports
/root    192.168.4.205(rw,no_root_squash)
/usr/src    192.168.4.0/24(ro)

[root@svr5 ~]# systemctl restart rpcbind      //启用RPC机制
[root@svr5 ~]# # systemctl restart nfs        //启用nfs服务
```

使用NFS客户端

- 查看NFS共享列表
  - showmount -e [服务器地址]

```
[root@pc205 ~]# showmount -e 192.168.4.5
Export list for 192.168.4.5:
/root    192.168.4.205
/usr/src 192.168.4.0/24
```

知识讲解

使用NFS客户端（续1）

- 挂载NFS共享目录
  - mount [-t nfs] 服务器地址:共享目录 挂载点

```
[root@pc205 ~]# service rpcbind restart                    //启用RPC机制
[root@pc205 ~]# mount 192.168.4.5:/root /mnt/root5/
[root@pc205 ~]# ls /mnt/root5/
anaconda-ks.cfg install.log install.log.syslog
.. ..
```

注意事项：
1. 从未授权的客户机将无法挂载此共享
2. 若未启用 no_root_squash，挂载后会无权限浏览（750）

知识讲解

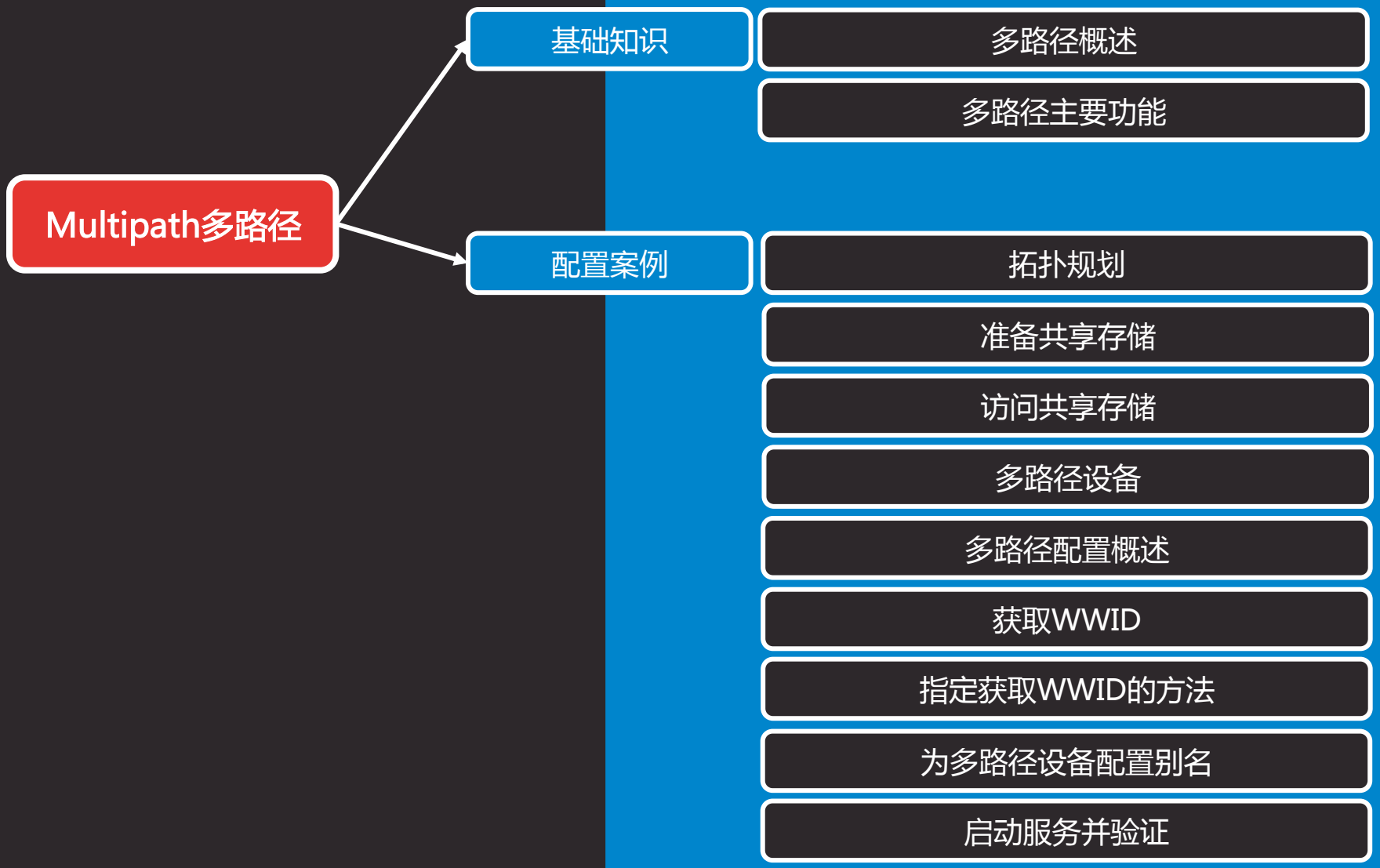
案例3：配置并访问NFS共享

课堂练习

1. 利用NFS发布2个共享目录
   - /root → 192.168.4.205，允许写入
   - /usr/src → 192.168.4.0/24网段，只读
2. 从客户机访问NFS共享
   - 查询/挂载上述NFS共享
   - 查看挂载点目录，并测试是否可写入
   - 为访问NFS共享目录 /usr/src 配置触发挂载

```
nfsdir      -fstype=nfs,ro      192.168.4.5:/usr/src
```

Multipath多路径



Tedu.cn
达内教育

基础知识

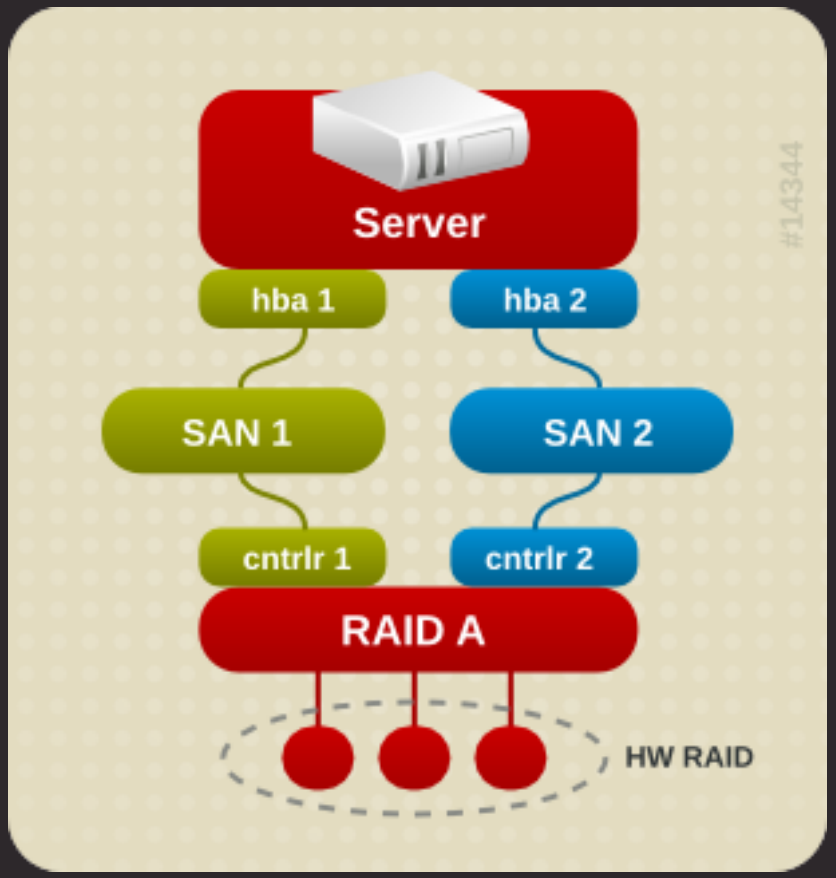
多路径概述

- 当服务器到某一存储设备有多条路径时，每条路径都会识别为一个单独的设备
- 多路径允许您将服务器节点和储存阵列间的多个I/O路径配置为一个单一设备
- 这些 I/O 路径是可包含独立电缆、交换器和控制器的实体 SAN 链接
- 多路径集合了 I/O 路径，并生成由这些集合路径组成的新设备

知识讲解

多路径主要功能

- 冗余
  - 主备模式，高可用
- 改进的性能
  - 主主模式，负载均衡



知识讲解

Tedu.cn
达内教育

配置案例

拓扑规划

- 利用iSCSI实现多路径

应用服务器	IP 地址
eth0	192.168.1.10/24
eth1	192.168.2.10/24

存储节点	IP 地址
eth0	192.168.1.20/24
eth1	192.168.2.20/24

知识讲解

准备共享存储

- 配置iSCSI服务端
  - 准备共享介质（分区、LV或磁盘镜像）
  - 安装scsi-target-utils
  - 准备规划iqn名称
  - 修改target.conf配置文件，以提供存储
  - 启动tgtd服务

知识讲解

访问共享存储

- 因为到达共享存储有两条路径，所以需要在两条路径上都执行发现命令

```
[root@node4 ~]# iscsiadm --mode discovery --type sendtargets \
--portal 192.168.1.20 --discover
[root@node4 ~]# iscsiadm --mode discovery --type sendtargets \
--portal 192.168.2.20 --discover
```

- 设置开机自启

```
[root@svr4 ~]# systemctl enable iscsi
[root@svr4 ~]# systemctl enable iscsid
```

知识讲解

多路径设备

- 若没有 DM Multipath，从服务器节点到储存控制器的每一条路径都会被系统视为独立的设备，即使 I/O 路径连接的是相同的服务器节点到相同的储存控制器也是如此
- DM Multipath 提供了有逻辑的管理 I/O 路径的方法，即在基础设备顶端生成单一多路径设备

知识讲解

知识讲解

多路径配置概述

Technology Tarena 达内科技

- 安装软件包

```
[root@node4 ~]# yum  install  -y  device-mapper-multipath
```

- 使用 mpathconf 命令创建配置文件并启用多路径

```
[root@node4 ~]# mpathconf  --user_friendly_names n
```

- 若无需编辑该配置文件，可使用此命令启动多路径守护程序

多路径设备识别符

- 每个多路径设备都有一个 WWID（全球识别符），它是全球唯一的、无法更改的号码
- 默认情况下会将多路径设备的名称设定为它的 WWID
- 可以在多路径配置文件中设置user_friendly_names选项，该选项可将别名设为格式为 mpathn 的节点唯一名称
- 也可以自定义存储设备名称

获取WWID

- 假如共享存储在本地被识别为/dev/sdb和/dev/sdc，那么获取它WWID的方法是：

```
[root@node4 ~]# /lib/udev/scsi_id --whitelisted --device=/dev/sdb
[root@node4 ~]# /lib/udev/scsi_id --whitelisted --device=/dev/sdc
```

- 因为两个设备虽然名称不一样，但是实际上是一个设备，所以他们的WWID是相同的

指定获取WWID的方法

- 在配置文件中声明获取WWID的方法

```
[root@node4 ~]# vim /etc/multipath.conf

defaults {
        user_friendly_names no
        getuid_callout      "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
}
```

知识讲解

为多路径设备配置别名

- 根据得到的WWID，为多路径设备配置别名

```
[root@node4 ~]# vim /etc/multipath.conf
# 在尾部添加以下内容：
multipaths {
        multipath {
                wwid  "1IET     00010001"
                alias    mpatha
        }
}
```

知识讲解

启动服务并验证

- 启动服务

```
[root@node4 ~] # systemctl start multipathd
[root@node4 ~] # systemctl enable multipathd
```

- 验证

```
[root@node4 ~]# ls  /dev/mapper     #mpatha即为多路径设备
[root@node4 ~]# multipath -rr       # 重新加载多径信息
[root@node4 ~]# multipath -ll       #查看多径信息
```

- 分区
  - 为/dev/mapper/mpatha分区，得到的第一个分区名为/dev/mapper/mpathap1

知识讲解

案例4：部署Multipath多路径环境

通过Multipath，实现以下目标

- 在共享存储服务器上配置iSCSI
- 应用服务器上配置iSCSI，发现远程共享存储
- 应用服务器上配置Multipath，将相同的共享存储映射为同一个名称

课堂练习

总结和答疑

客户端检测不到磁盘

问题现象

- iSCSI客户端可以发现target
- 登陆之后，本地并没有出现新的磁盘设备

知识讲解

故障分析及排除

- 原因分析
  - target端没有出现LUN1
  - 对target端进行修改，重启服务后，target没有LUN1
- 解决办法
  - 只能重启操作系统

知识讲解

NFS无法写入数据

问题现象

- NFS服务器端已配置共享输出目录为读写权限
- 客户端可以正常挂载服务器的输出目录
- 客户端向服务器输出目录写入数据失败

知识讲解

故障分析及排除

- 原因分析
  - 服务器没有配置本地的写权限
  - 输出目录权限的最终权限，由本地权限和配置文件内的权限共同决定
- 解决办法
  - 将目录的本地权限改为777

知识讲解