

CLUSTER DAY02



# 集群与存储

**NSD CLUSTER**

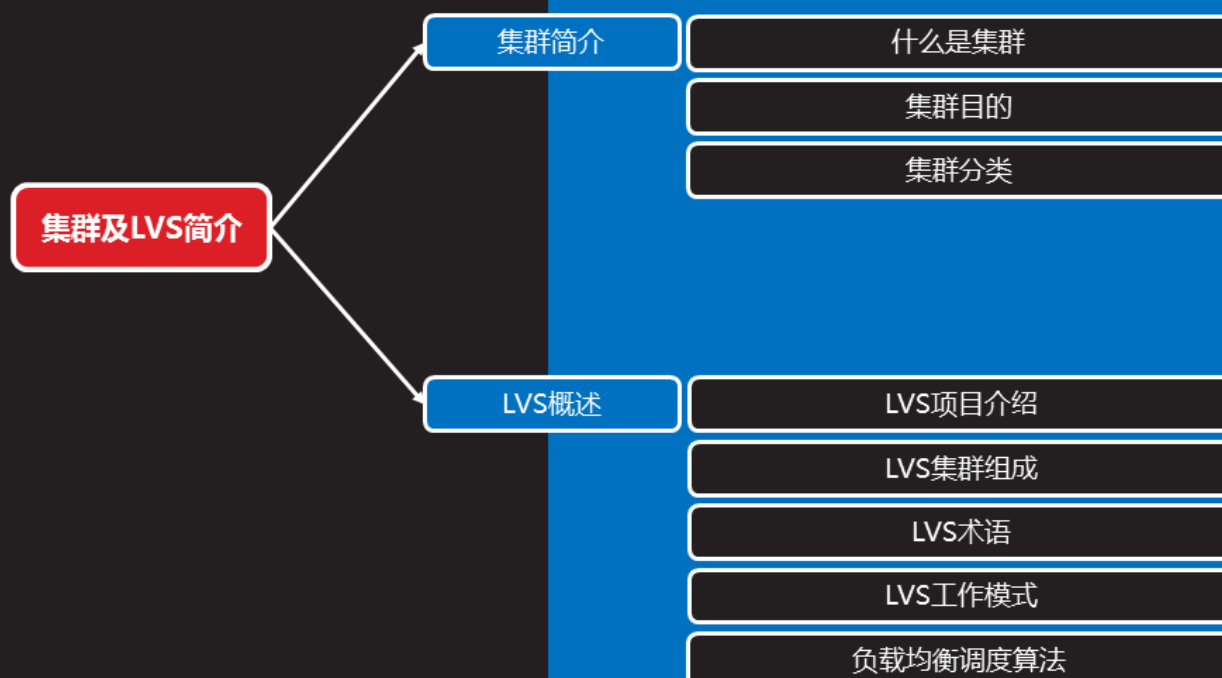
**DAY02**

# 内容

上午	09:00 ~ 09:30	作业讲解和回顾
	09:30 ~ 10:20	集群及LVS简介
	10:30 ~ 11:20	
	11:30 ~ 12:20	LVS-NAT集群
下午	14:00 ~ 14:50	
	15:00 ~ 15:50	LVS-DR集群
	16:10 ~ 17:00	
	17:10 ~ 18:00	总结和答疑



## 集群及LVS简介



# 集群简介

## 什么是集群

知识讲解

- 一组通过高速网络互联的计算组，并以单一系统的模式加以管理
- 将很多服务器集中起来一起，提供同一种服务，在客户端看来就象是只有一个服务器
- 可以在付出较低成本的情况下获得在性能、可靠性、灵活性方面的相对较高的收益
- 任务调度是集群系统中的核心技术



## 集群目的

知识讲解

- 提高性能
  - 如计算密集型应用，如：天气预报、核试验模拟
- 降低成本
  - 相对百万美元级的超级计算机，价格便宜
- 提高可扩展性
  - 只要增加集群节点即可
- 增强可靠性
  - 多个节点完成相同功能，避免单点失败



## 集群分类

知识讲解

- 高性能计算集群HPC
  - 通过以集群开发的并行应用程序，解决复杂的科学问题
- 负载均衡（LB）集群
  - 客户端负载在计算机集群中尽可能平均分摊
- 高可用（HA）集群
  - 避免单点故障，当一个系统发生故障时，可以快速迁移



# LVS概述



## LVS项目介绍

知识讲解

- Linux 虚拟服务器 ( LVS ) 是章文嵩在国防科技大学就读博士期间创建的
- LVS可以实现高可用的、可伸缩的Web、Mail、Cache和Media等网络服务
- 最终目标是利用Linux操作系统和LVS集群软件实现一个高可用、高性能、低成本的服务器应用集群



## LVS集群组成

知识讲解

- 前端：负载均衡层
  - 由一台或多台负载调度器构成
- 中间：服务器群组层
  - 由一组实际运行应用服务的服务器组成
- 底端：数据共享存储层
  - 提供共享存储空间的存储区域



## LVS术语

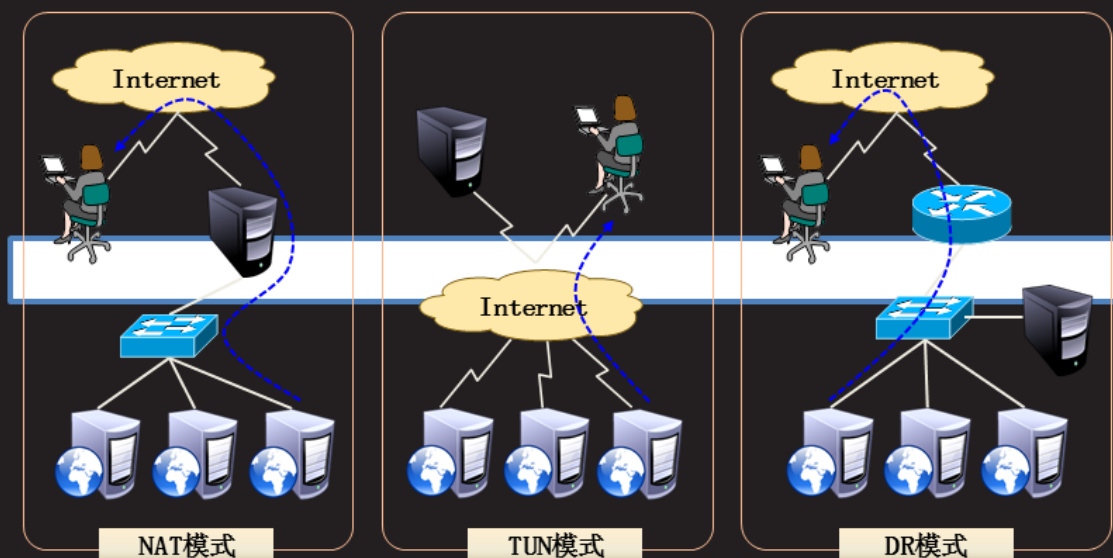
知识讲解

- Director Server：调度服务器
  - 将负载分发到Real Server的服务器
- Real Server：真实服务器
  - 真正提供应用服务的服务器
- VIP：虚拟IP地址
  - 公布给用户访问的虚拟IP地址
- RIP：真实IP地址
  - 集群节点上使用的IP地址
- DIP：调度器连接节点服务器的IP地址



# LVS工作模式

知识讲解



## LVS工作模式（续1）

知识讲解

- VS/NAT
  - 通过网络地址转换实现的虚拟服务器
  - 大并发访问时，调度器的性能成为瓶颈
- VS/DR
  - 直接使用路由技术实现虚拟服务器
  - 节点服务器需要配置VIP，注意MAC地址广播
- VS/TUN
  - 通过隧道方式实现虚拟服务器



# 负载均衡调度算法

知识讲解

- LVS目前实现了10种调度算法
- 常用调度算法有4种
  - 轮询 ( Round Robin )
  - 加权轮询 ( Weighted Round Robin )
  - 最少连接 ( Least Connections )
  - 加权最少连接 ( Weighted Least Connections )



## 负载均衡调度算法 ( 续1 )

知识讲解

- 轮询 ( Round Robin )
  - 将客户端请求平均分发到Real Server
- 加权轮询 ( Weighted Round Robin )
  - 根据Real Server权重值进行轮询调度
- 最少连接 ( Least Connections )
  - 选择连接数最少的服务器





## 负载均衡调度算法（续2）

知识讲解

- 加权最少连接
  - 根据Real Server权重值，选择连接数最少的服务器
- 源地址散列（Source Hashing）
  - 根据请求的目标IP地址，作为散列键(Hash Key)从静态分配的散列表找出对应的服务器



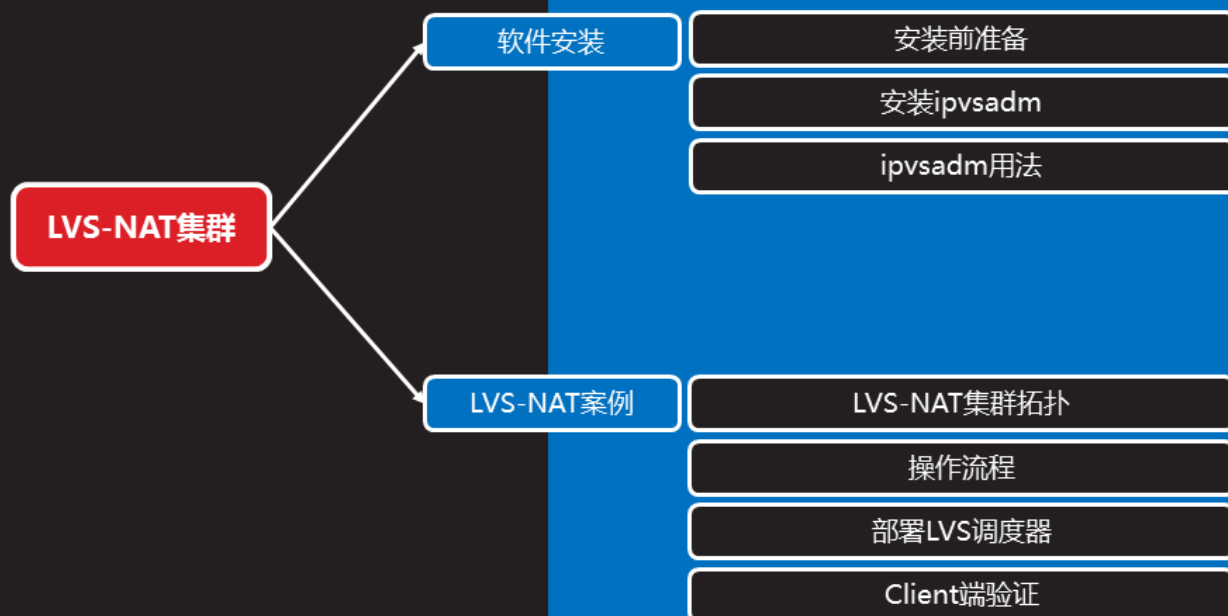
## 负载均衡调度算法（续3）

知识讲解

- 其他调度算法
  - 基于局部性的最少链接
  - 带复制的基于局部性最少链接
  - 目标地址散列（Destination Hashing）
  - 最短的期望的延迟
  - 最少队列调度



# LVS-NAT集群



## 安装软件

## 安装前准备

- LVS的IP负载均衡技术是通过IPVS模块实现的
- IPVS模块已成为Linux组成部分

知识讲解

```
[root@svr1 ~]# grep -i 'ipvs' /boot/config-3.10.0-327.el7.x86_64
# IPVS transport protocol load balancing support
# IPVS scheduler
# IPVS SH scheduler
# IPVS application helper
```



## 安装ipvsadm

- 使用rpm命令安装ipvsadm

知识讲解

```
[root@svr1 Packages]# rpm -ihv ipvsadm-1.27-7.el7.x86_64.rpm.rpm
warning: ipvsadm-1.27-7.el7.x86_64.rpm.rpm: Header V3 RSA/SHA256
Signature, key ID fd431d51: NOKEY
Preparing... ##### [100%]
 1:ipvsadm      ##### [100%]
[root@svr1 Packages]#
[root@svr1 Packages]# ipvsadm -v
ipvsadm v1.27 2008/5/15 (compiled with popt and IPVS v1.2.1)
```



# ipvsadm用法

知识讲解

- 创建虚拟服务器

- -A 添加虚拟服务器
- -t 设置群集地址 ( VIP , Virtual IP )
- -s 指定负载调度算法

```
[root@svr1 ~]# ipvsadm -A -t 172.16.16.172:80 -s rr
```



# ipvsadm用法 ( 续1 )

知识讲解

- 添加、删除服务器节点

- -a 添加真实服务器
- -d 删除真实服务器
- -r 指定真实服务器 ( Real Server ) 的地址
- -m 使用NAT模式 ; -g、-i分别对应DR、TUN模式
- -w 为节点服务器设置权重 , 默认为1

```
[root@svr1 ~]# ipvsadm -a -t 172.16.16.172:80 -r 192.168.7.21:80 -m
```

```
[root@svr1 ~]# ipvsadm -d -r 192.168.7.24:80 -t 172.16.16.172:80
```



## ipvsadm用法 ( 续2 )

- 查看IPVS

知识讲解

```
[root@svr1 ~]# ipvsadm -Ln
TCP 172.16.16.172:80 rr
-> 192.168.7.21:80      Masq  1   2   7
-> 192.168.7.22:80      Masq  1   3   9
-> 192.168.7.23:80      Masq  1   2   8
-> 192.168.7.24:80      Masq  1   4   6
```



## 案例1：ipvsadm命令用法

课堂练习

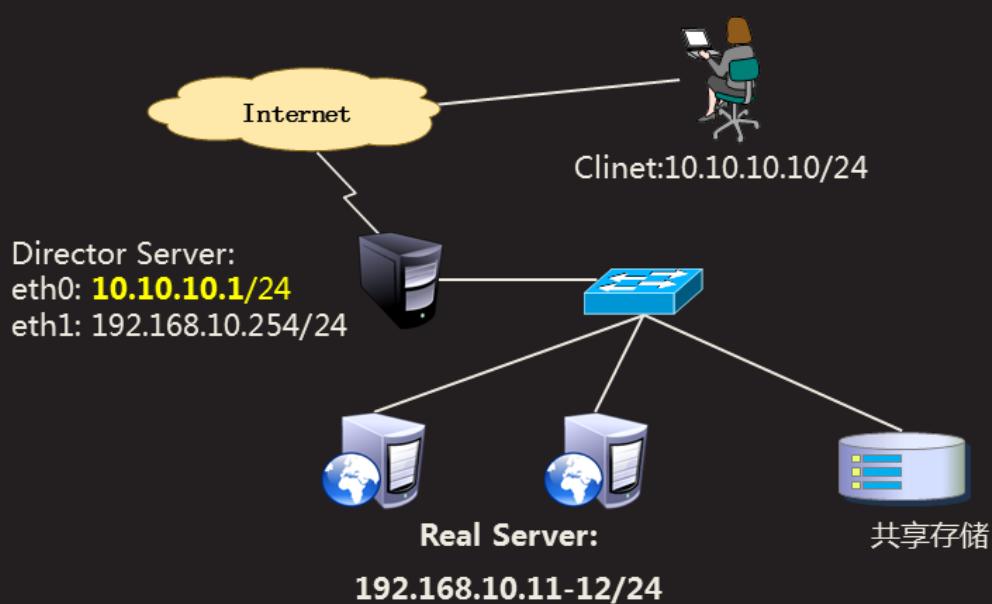
- 使用命令添加基于TCP一些的集群服务
- 在集群中添加若干台后端真实服务器
- 实现同一客户端访问，调度器分配固定服务器
- 会使用ipvsadm实现规则的增、删、改
- 保存ipvsadm规则



# LVS-NAT案例

## LVS-NAT集群拓扑

知识讲解



## 操作流程

知识讲解

- Real Server :
  - 配置WEB服务器
- Director Server :
  - 在上安装并启用ipvsadm
  - 创建虚拟服务器
  - 向虚拟服务器中加入节点
- Client :
  - 连接虚拟服务器测试



## 部署LVS调度器

知识讲解

- 打开ip\_forward

```
[root@svr1 ~]# vim /etc/sysctl.conf
net.ipv4.ip_forward = 1
[root@svr1 ~]# sysctl -p
```
- 启动ipvsadm

```
[root@svr1 ~]# systemctl start ipvsadm
[root@svr1 ~]# systemctl enable ipvsadm
```



## 部署LVS调度器（续1）

知识讲解

- 创建虚拟服务器，VIP为10.10.10.1，采用的调度算法为Round Robin

```
[root@svr1 ~]# ipvsadm -A -t 10.10.10.1:80 -s rr
```

- 向虚拟服务器中加入节点，并指定权重分别为1和2，目前权重不起作用（为什么？）

```
[root@svr1 ~]# ipvsadm -a -t 10.10.10.1:80 -r 192.168.10.11 -m -w 1
```

```
[root@svr1 ~]# ipvsadm -a -t 10.10.10.1:80 -r 192.168.10.12 -m -w 2
```



## 部署LVS调度器（续2）

知识讲解

- 查看配置

```
[root@svr1 ~]# ipvsadm -L -n
```

//注意：这两个选项不写反

- 保存配置

```
[root@svr1 ~]# ipvsadm -s > /etc/sysconfig/ipvsadm
```

- 修改Director调度算法为WRR

```
[root@svr1 ~]# ipvsadm -E -t 10.10.10.1:80 -s wrr
```





# Client端验证

知识讲解

- 通过web浏览器访问
- 使用ab进行大并发测试

```
[root@client ~]# ab -c 10 -n 1000 http://10.10.10.1/index.html
```

- 在Director上查看连接数

```
[root@svr1 ~]# ipvsadm -L -n
```



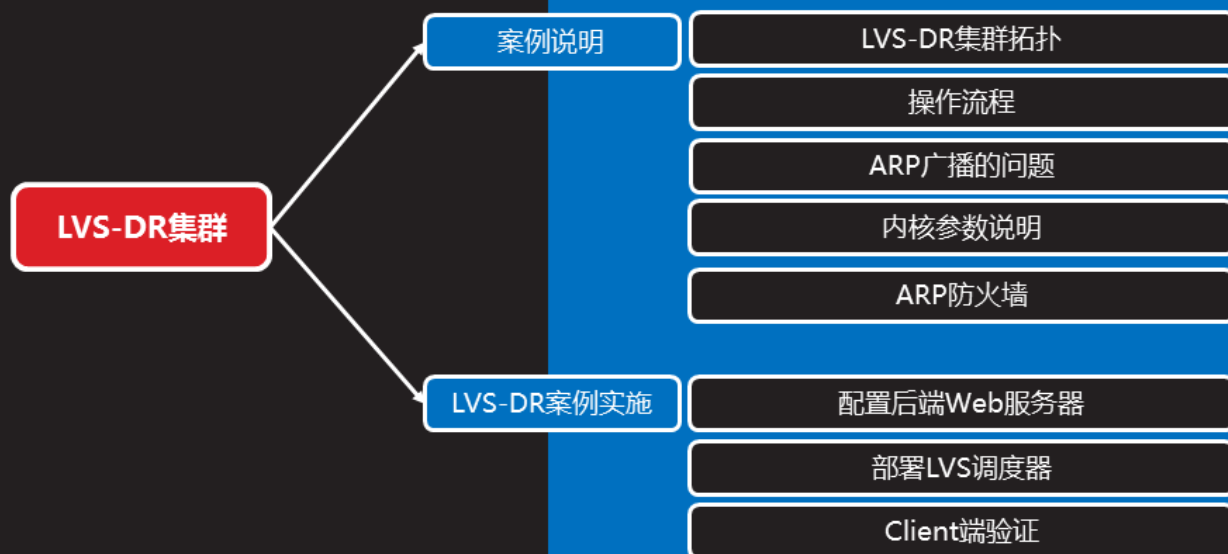
## 案例2：部署LVS-NAT集群

课堂练习

- 准备三台虚拟机
  - 1台作为Director
  - 2台做为Real Server
  - 物理机作为客户端
- Real Server安装httpd服务
  - 每台服务器的主页内容不同
- 在Director上安装双网卡，并将其配置为NAT模式
- 集群使用加权轮询调度算法，真实服务器权重分别为1和2



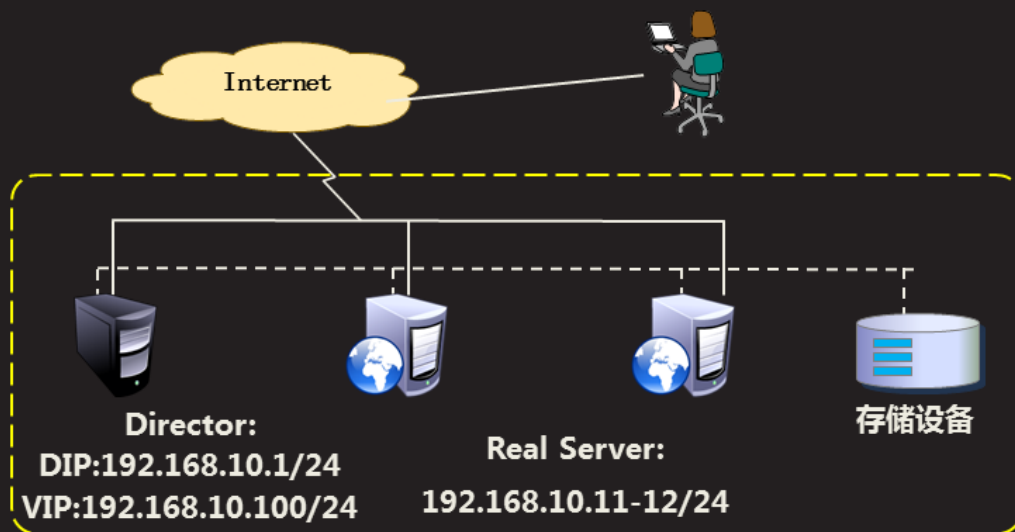
# LVS-DR集群



## 案例说明

# LVS-DR集群拓扑

知识讲解



## 操作流程

知识讲解

- Real Server :
  - 配置WEB服务器
  - 配置辅助IP地址、调整内核参数
- Director Server :
  - 在上安装并启用ipvsadm
  - 配置辅助IP地址
  - 创建虚拟服务器、向虚拟服务器中加入节点
- Client :
  - 连接虚拟服务器测试



# ARP广播的问题

知识讲解

- 当客户端发起访问VIP 对应的域名的请求时，根据网络通信原理会产生ARP 广播
- 因为负载均衡器和真实的服务器在同一网络并且VIP 设置在集群中的每个节点上
- 此时集群内的真实服务器会尝试回答来自客户端的ARP广播，这就会产生问题，大家都说我是"VIP"



# 内核参数说明

知识讲解

- arp\_ignore ( 定义回复ARP广播的方式 )
  - 0(默认值)  
回应所有的本地地址ARP广播，本地地址可以配置在任意网络接口
  - 1  
只回应配置在入站网卡接口上的任意IP地址的ARP广播



## 内核参数说明（续1）

知识讲解

- arp\_announce

- 0 (默认)

使用配置在任意网卡接口上的本地IP地址

- 2

对查询目标使用最适当的本地地址。在此模式下将忽略这个IP数据包的源地址并尝试选择与能与该地址通信的本地地址。首要是选择所有的网络接口的子网中外出访问子网中包含该目标IP地址的本地地址。如果没有合适的地址被发现，将选择当前的发送网络接口或其他的有可能接受到该ARP回应的网络接口来进行发送



## ARP防火墙

- 使用ARP防火墙也可以禁止对VIP的ARP请求

```
[root@client ~]# yum -y install arptables_jf
```

```
[root@client ~]# arptables -A IN -d <virtual_ip> -j DROP
```

```
[root@client ~]# arptables -A OUT -s <virtual_ip> -j mangle \> --mangle-ip-s <real_ip>
```

知识讲解



# LVS-DR案例实施

## 配置后端Web服务器

- 配置辅助VIP地址

```
[root@web1 ~]# ifconfig lo:0 192.168.10.100 netmask \  
> 255.255.255.255 broadcast 192.168.10.100 up
```

- 调整内核参数

```
[root@web1 ~]# echo 1 > /proc/sys/net/ipv4/conf/lo/arp_ignore  
[root@web1 ~]# echo 2 > /proc/sys/net/ipv4/conf/lo/arp_announce  
[root@web1 ~]# echo 1 > /proc/sys/net/ipv4/conf/all/arp_ignore  
[root@web1 ~]# echo 2 > /proc/sys/net/ipv4/conf/all/arp_announce
```

## 配置LVS调度器

知识讲解

- 配置辅助IP地址

```
[root@svr1 ~]# ifconfig eth0:1 192.168.10.100 broadcast 192.168.10.100\  
> netmask 255.255.255.255 up
```

- 创建虚拟服务器

```
[root@svr1 ~]# ipvsadm -A -t 192.168.10.100:80 -s wlc
```

- 向虚拟服务器中加入节点

```
[root@svr1 ~]# ipvsadm -a -t 192.168.10.100:80 -g -r 192.168.10.11 -w 1  
[root@svr1 ~]# ipvsadm -a -t 192.168.10.100:80 -g -r 192.168.10.12 -w 2
```



## Client端验证

知识讲解

- 通过web浏览器访问
- 使用ab进行大并发测试

```
[root@client ~]# ab -c 10 -n 1000 http://10.10.10.1/index.html
```

- 在Director上查看连接数

```
[root@svr1 ~]# ipvsadm -L -n
```



## 案例3：部署LVS-DR集群

课堂练习

- 使用LVS实现DR模式的集群调度服务器，为用户提供Web服务：
  - 客户端IP地址为192.168.4.10
  - VS调度器VIP地址为192.168.4.15
  - LVS调度器DIP地址设置为192.168.4.5
  - 真实Web服务器地址分别为192.168.4.100、192.168.4.200
  - 使用加权轮询调度算法，web1的权重为1，web2的权重为2



## 总结和答疑



## 总结和答疑



## 加权轮询不生效

## 问题现象

知识讲解

- 在LVS调度器上配置的real server权重不一样
- 客户机访问服务器时，real server的权重不起作用，只是简单的轮询



## 故障分析及排除

知识讲解

- 原因分析
  - LVS调度器虽然为real server设置了权重，但是调度算法仍然是轮询
  - 轮询算法根本不检查权重值
- 解决办法
  - 修改调度算法为wrr



# 服务无法启动

## 问题现象

- 启动LVS服务时，启动失败
- 尝试以restart方式启动服务，仍然失败

知识讲解



# 故障分析及排除

## 知识讲解

- 原因分析
  - 检查服务状态和日志，提示没有/etc/sysconfig/ipvsadm文件
  - 该文件是调度规则文件
- 解决办法
  - 先将规则存盘，再重启服务

