2018年8月6日

purple passion

LVS: linux virtual service(内核实现,效率高,iptables也是内核实现),主要实现4层传输层的负载均衡服务,如httpd等.

四层七层区别

四层通过虚拟 IP + 端口接收请求，然后再分配到真实的服务器；

七层通过虚拟的 URL 或主机名接收请求，然后再分配到真实的服务器。

集群技术

集群：1.将众多服务器集中起来一起，提供一种服务，在客户端看来就象是只有一个服务器。2.可以在付出较低成本情况下获得性能、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

提高性能(天气预报)、降低成本（相对百万美元级computer）、提高可扩展性（可增加集群节点）、增强可靠性（避免单点故障）

服务器资源:CPU,内存,网络,硬盘I/O

HPC高性能计算,为了解决计算量

LB负载均衡,为解决并发量

HA高可用,为了解决单点故障

(cip) (vip) (dip) (rip)

client ----------->调度器(director server) -->应用服务器(real server)--->storage 后三段为集群,

LVS工作模式(三种)

第一种:NAT,中间的调度器主机是瓶颈,最多只能10-20个real server

第二种:DR(直接路由模式),只服务用户端的请求,调用的read server后直接从路由回馈给客户端,能到100个read server(用最多是这模式)

第三种:TUN,通过隧道回应,相当VPN的隧道,类似直接路由.这模式用得不多

第四种:fullnat

++++++++++++++++++++++++++++++++++++++++++++

1台client

1台director server

2台real server

客户端:CIP(eth2):201.1.1.200

调度器:VIP(eth2):201.1.1.100和DIP(eth0):192.168.4.100

real server(两台):RIP(eth0):192.168.4.11/12 GW:192.168.4.100

+++++++++++++++++++++++++++++++++++++++++++++++++

一.LVS-NAT调度器

1.配置调度器 VIP主机

装包ipvsadm

vip ~]# yum -y install ipvsadm

创建虚拟服务器-A添加虚拟服务器,-t设置群集地址virtual IP,-s指定负载调度算法

vip ~]# ipvsadm -L -n

(添加虚拟服务器-A,-t添加vip地址201.1.1.100)

vip ~]# ipvsadm -A -t 201.1.1.100:80 -s rr (-s rr是轮询算法,-s是指定负载调度算法,默认算法加权最少连接!!! 下面的-r是真实主机地址)

(添加真实服务器-a)

vip ~]# ipvsadm -a -t 201.1.1.100:80 -r 192.168.4.11 -m (-m是NAT模式,-g是DR模式,-i是TUN模式,-w设置权重值)

vip ~]# ipvsadm -a -t 201.1.1.100:80 -r 192.168.4.12 -m

vip ~]# ipvsadm -ln (ipvsadm -L -n也一样)

TCP 201.1.1.100:80 rr

-> 192.168.4.11:80 Masq 1 0 0

-> 192.168.4.12:80 Masq 1 0 0

**开启路由转发功能**

**(临时**)

[root@vip ~]# vim /proc/sys/net/ipv4/ip\_forward

1 (1为开启路由转发功能,在这里设置是临时的)

**(永久)**

vip ~]# vim /etc/sysctl.conf

net.ipv4.ip\_forward = 1

vip ~]# sysctl -p

2.配置两台real server

安装apache

rip11 ~]#echo "re1" > /var/www/html/index.html

rip12 ~]#echo "re2" > /var/www/html/index.html

3.客户端测试

[root@cip ~]# curl 201.1.1.100

rs2

[root@cip ~]# curl 201.1.1.100

rs1

(如果在VIP主机测试没问题,在CIP客户机测试卡死不显示,那么就是res1和2主机没配网关或配错)

捉包看转发过程

vip ~]# tcpdump -nn port 80

cip ~]# curl 201.1.1.100 再去看捉包的终端

cip ~]# tcpdump -i eth2 -nn port 80 (-i指定网卡端口)

wrr算法

vip ~]# ipvsadm -C (-C清空规则)

vip ~]# ipvsadm -A -t 201.1.1.100:80 -s wrr (wrr顺序不能变)

vip ~]# ipvsadm -a -t 201.1.1.100:80 -r 192.168.4.11 -m -w2

vip ~]# ipvsadm -a -t 201.1.1.100:80 -r 192.168.4.12 -m -w1

vip ~]# ipvsadm -nl

验证

cip ~]# curl 201.1.1.100

rs2

cip ~]# curl 201.1.1.100

rs1

cip ~]# curl 201.1.1.100

rs1

开机自动运行ipvsadm规则

vip ~]#ipvsadm -save -n > /etc/sysconfig/ipvsadm

vip ~]#systemctl start ipvsadm

++++++++++++++++++++++++++++++++++++++++++++++++

**二.LVS-DR调度器**

查看ARP缓存表

vip ~]# arp -n

192.168.4.12 ether 52:54:00:68:19:01 C eth0

192.168.4.254 ether 52:54:00:37:78:11 C eth0

192.168.4.11 ether 52:54:00:cb:9c:22 C eth0

201.1.1.200 ether 52:54:00:ee:f7:6a C eth2

201.1.1.254 ether 52:54:00:88:2b:9c C

vip ~]# arp -d 192.168.4.11 (删除192.168.4.11ip的mac地址)

vip ~]# ping 192.168.4.11 再查看arp -n 又有192.168.4.11的mac地址

ip命令用法

ip link help查看使用帮助

]#ip a 或 ip address 查看所有ip地址

rip11 ~]# ip addr ls dev eth0

rip11 ~]# ip addr add dev eth0 192.168.4.11/24 增加ip地址

rip11 ~]# ip addr ls dev eth0

inet 192.168.4.11/24 brd 192.168.4.255 scope global eth0

valid\_lft forever preferred\_lft forever

inet 192.168.5.11/24 scope global eth0

rip11 ~]# ip add del dev eth0 192.168.5.11/24 删除ip地址

添加一个虚拟的网络设备

rip11 ~]# ip link add veth-a type veth peer veth-b //添加时是添加一对的

桥就是交换机

临时配置

rip11 ~]# ip a a dev veth-a 192.168.4.101/24

vip ~]# ping 192.168.4.101 能ping通

vip ~]# arp -n

Iface

192.168.4.11 ether 52:54:00:cb:9c:22 C eth0

192.168.4.101 ether 52:54:00:cb:9c:22

(原因在rip11上老是4.11帮4.101收发)

取消以上的帮应答

rip11 ~]# echo 1 > /proc/sys/net/ipv4/conf/eth0/arp\_ignore

rip11 ~]# echo 2 > /proc/sys/net/ipv4/conf/eth0/arp\_announce

删除arp缓存表条目

vip ~]# arp -d 192.168.4.101 提前删除绶存有效期

vip ~]# ping 192.168.4.101 ping不通了

sip:cip ---> des ip:vip des ip:cip <----- sip:rip (这样是不合理的包)

+++++++++++++++++++++++++++++++++++++++

201.1.1.200<-----------------201.1.1.101

<----------------------------------------------------------|

|------>real11 201.1.1.102

client ---------->VIP------------->|

201.1.1.200 vip:201.1.1.100 |------>real12 201.1.1.103

dip:201.1.1.101

<------------------------------------------------------------|

201.1.1.200<-----------------201.1.1.101

两台real servser主机做以下配置

rip11 ~]# nmcli con add type ethernet ifname eth2 con-name eth2

rip11 ~]# nmcli con modify eth2 ipv4.method manual ipv4.addresses 201.1.1.102/24 connection.autoconnect yes

rip11 ~]# nmcli con up eth2

配置vip201.1.1.101

vip ~]# ip a ls dev eth2

vip ~]# ip a a dev eth2 201.1.1.101/24

配虚拟服务器

vip ~]# ipvsadm -A -t 201.1.1.101:80 -s rr

配置真实服务器

vip ~]# ipvsadm -a -t 201.1.1.101:80 -r 201.1.1.102 -g

vip ~]# ipvsadm -a -t 201.1.1.101:80 -r 201.1.1.103 -g

配置real server主机

1.在lo设备上配置vip

rip11 ~]# ip a a dev lo 201.1.1.101/32 (32或默认这个网络中只有一个,lo只有自已ping通自已)

2.禁用arp请求

把CIP网关配成201.1.2.254,VIP和RIP11,RIP12的网关都配成201.1.1.254

在rip11和rip12主机上做同样操作

rip11 ~]# echo 1 > /proc/sys/net/ipv4/conf/all/arp\_ignore

rip11 ~]# echo 2 > /proc/sys/net/ipv4/conf/all/arp\_announce (默认都是0) //这两步是两台real server主机返回IP为201.1.1.101的关建

验证结果

[root@cip ~]# curl 201.1.1.101

rs2

[root@cip ~]# curl 201.1.1.101

rs1

问题:

C:curl 201.1.1.101 --->全是rs2

R12:echo 0和2没设置 R11设了

V:没配临时 eth2 201.1.1.101 R11,12配了

先把R12 echo 0和2配好,在C:curl 201.1.1.101 :没有主机路由

再配V:没配临时 eth2 201.1.1.101 就可以轮询访问了

++++++++++++++++++++++++++++++++++++++

再配个路由器,真机作为路由器,

client --201.1.2.254(route)201.1.1.254 ----lb

[root@room9pc01 ~]# cat /proc/sys/net/ipv4/ip\_forward

1

cip ~]# nmcli con add type ethernet ifname eth3 con-name eth3

cip ~]# nmcli con modify eth3 ipv4.method manual ipv4.addresses 201.1.2.200/24 ipv4.gateway 201.1.2.254 connection.autoconnect yes

cip ~]# nmcli con up eth3

在real server主机上添加网关

rip11 ~]# nmcli con modify eth2 ipv4.method manual ipv4.gateway 201.1.1.254

rip11 ~]# nmcli con up eth2

删除网关

rip11 ~]# nmcli con modify eth0 ipv4.method manual ipv4.gateway 0.0.0.0

rip11 ~]# nmcli con up eth0

rip11 ~]# route -n

Kernel IP routing table

Destination Gateway Genmask Flags Metric Ref Use Iface

0.0.0.0 201.1.1.254 0.0.0.0 UG 100 0 0 eth2

192.168.4.0 0.0.0.0 255.255.255.0

[root@room9pc01 ~]# iptables -nL 查看

[root@room9pc01 ~]# iptables -F -A FORWARD 清空

刚才配置VIP主机网关时,激活eth2网卡时把临时配置的201.1.1.101更新掉了,需再配置

vip ~]# ip a a dev eth2 201.1.1.101/24

临时删除客户端上的网关

cip ~]# route del default

[root@room9pc01 ~]# iptables -t nat -A PREROUTING -d 201.1.2.254 -p tcp --dport 80 -j DNAT --to 201.1.1.101

[root@room9pc01 ~]# iptables -t nat -A POSTROUTING -s 201.1.1.0/24 -j SNAT --to 201.1.2.254

查看iptables的nat规则

[root@room9pc01 ~]# iptables -t nat -nL

防问201.1.2.254就是去访问201.1.1.101

[root@cip ~]# curl 201.1.2.254

rs2

[root@cip ~]# curl 201.1.2.254

rs1