LVS

1. LVS简介

LVS是Linux Virtual Server的简称，也就是Linux虚拟服务器, 是一个由章文嵩博士发起的自由软件项目，它的官方站点是[www.linuxvirtualserver.org](http://www.linuxvirtualserver.org/)。现在LVS已经是 Linux标准内核的一部分，在Linux2.4内核以前，使用LVS时必须要重新编译内核以支持LVS功能模块，但是从Linux2.4内核以后，已经完全内置了LVS的各个功能模块，无需给内核打任何补丁，可以直接使用LVS提供的各种功能。

使用LVS技术要达到的目标是：通过LVS提供的负载均衡技术和Linux操作系统实现一个高性能、高可用的服务器群集，它具有良好可靠性、可扩展性和可操作性。从而以低廉的成本实现最优的服务性能。

LVS自从1998年开始，发展到现在已经是一个比较成熟的技术项目了。可以利用LVS技术实现高可伸缩的、高可用的网络服务，例如WWW服务、Cache服务、DNS服务、FTP服务、MAIL服务、视频/音频点播服务等等，有许多比较著名网站和组织都在使用LVS架设的集群系统，例如：Linux的门户网站（[www.linux.com](http://www.linux.com/)）、向RealPlayer提供音频视频服务而闻名的Real公司（[www.real.com](http://www.real.com/)）、全球最大的开源网站（sourceforge.net）等。  
二、 LVS体系结构

使用LVS架设的服务器集群系统有三个部分组成：最前端的负载均衡层，用Load Balancer表示，中间的服务器群组层，用Server Array表示，最底端的数据共享存储层，用Shared Storage表示，在用户看来，所有的内部应用都是透明的，用户只是在使用一个虚拟服务器提供的高性能服务。

LVS体系结构如图1所示：

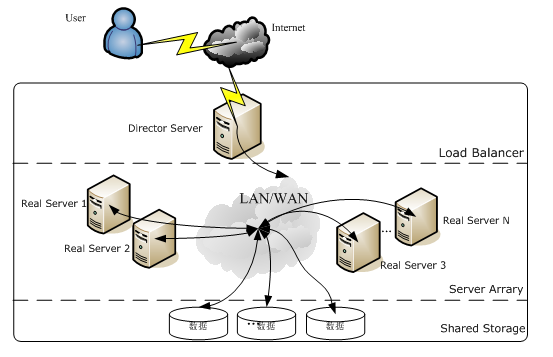
[](http://img1.51cto.com/attachment/201104/110349553.png)

图1 LVS的体系结构

下面对LVS的各个组成部分进行详细介绍：  
 Load Balancer层：位于整个集群系统的最前端，有一台或者多台负载调度器（Director Server）组成，LVS模块就安装在Director Server上，而Director的主要作用类似于一个路由器，它含有完成LVS功能所设定的路由表，通过这些路由表把用户的请求分发给Server Array层的应用服务器（Real Server）上。同时，在Director Server上还要安装对Real Server服务的监控模块Ldirectord，此模块用于监测各个Real Server服务的健康状况。在Real Server不可用时把它从LVS路由表中剔除，恢复时重新加入。  
 Server Array层：由一组实际运行应用服务的机器组成，Real Server可以是WEB服务器、MAIL服务器、FTP服务器、DNS服务器、视频服务器中的一个或者多个，每个Real Server之间通过高速的LAN或分布在各地的WAN相连接。在实际的应用中，Director Server也可以同时兼任Real Server的角色。  
 Shared Storage层：是为所有Real Server提供共享存储空间和内容一致性的存储区域，在物理上，一般有磁盘阵列设备组成，为了提供内容的一致性，一般可以通过NFS网络文件系统共享数据，但是NFS在繁忙的业务系统中，性能并不是很好，此时可以采用集群文件系统，例如Red hat的GFS文件系统，oracle提供的OCFS2文件系统等。

从整个LVS结构可以看出，Director Server是整个LVS的核心，目前，用于Director Server的操作系统只能是Linux和FreeBSD，linux2.6内核不用任何设置就可以支持LVS功能，而FreeBSD作为Director Server的应用还不是很多，性能也不是很好。

对于Real Server，几乎可以是所有的系统平台，Linux、windows、Solaris、AIX、BSD系列都能很好的支持。

三、  LVS集群的特点  
3.1  IP负载均衡与负载调度算法

1．IP负载均衡技术

负载均衡技术有很多实现方案，有基于DNS域名轮流解析的方法、有基于客户端调度访问的方法、有基于应用层系统负载的调度方法，还有基于IP地址的调度方法，在这些负载调度算法中，执行效率最高的是IP负载均衡技术。

LVS的IP负载均衡技术是通过IPVS模块来实现的，IPVS是LVS集群系统的核心软件，它的主要作用是：安装在Director Server上，同时在Director Server上虚拟出一个IP地址，用户必须通过这个虚拟的IP地址访问服务。这个虚拟IP一般称为LVS的VIP，即Virtual IP。访问的请求首先经过VIP到达负载调度器，然后由负载调度器从Real Server列表中选取一个服务节点响应用户的请求。

当用户的请求到达负载调度器后，调度器如何将请求发送到提供服务的Real Server节点，而Real Server节点如何返回数据给用户，是IPVS实现的重点技术，IPVS实现负载均衡机制有三种，分别是NAT、TUN和DR，详述如下：

VS/NAT： 即（Virtual Server via Network Address Translation）

也就是网络地址翻译技术实现虚拟服务器，当用户请求到达调度器时，调度器将请求报文的目标地址（即虚拟IP地址）改写成选定的Real Server地址，同时报文的目标端口也改成选定的Real Server的相应端口，最后将报文请求发送到选定的Real Server。在服务器端得到数据后，Real Server返回数据给用户时，需要再次经过负载调度器将报文的源地址和源端口改成虚拟IP地址和相应端口，然后把数据发送给用户，完成整个负载调度过程。

可以看出，在NAT方式下，用户请求和响应报文都必须经过Director Server地址重写，当用户请求越来越多时，调度器的处理能力将称为瓶颈。  
 VS/TUN ：即（Virtual Server via IP Tunneling）

也就是IP隧道技术实现虚拟服务器。它的连接调度和管理与VS/NAT方式一样，只是它的报文转发方法不同，VS/TUN方式中，调度器采用IP隧道技术将用户请求转发到某个Real Server，而这个Real Server将直接响应用户的请求，不再经过前端调度器，此外，对Real Server的地域位置没有要求，可以和Director Server位于同一个网段，也可以是独立的一个网络。因此，在TUN方式中，调度器将只处理用户的报文请求，集群系统的吞吐量大大提高。

VS/DR： 即（Virtual Server via Direct Routing）

也就是用直接路由技术实现虚拟服务器。它的连接调度和管理与VS/NAT和VS/TUN中的一样，但它的报文转发方法又有不同，VS/DR通过改写请求报文的MAC地址，将请求发送到Real Server，而Real Server将响应直接返回给客户，免去了VS/TUN中的IP隧道开销。这种方式是三种负载调度机制中性能最高最好的，但是必须要求Director Server与Real Server都有一块网卡连在同一物理网段上。

2．负载调度算法

上面我们谈到，负载调度器是根据各个服务器的负载情况，动态地选择一台Real Server响应用户请求，那么动态选择是如何实现呢，其实也就是我们这里要说的负载调度算法，根据不同的网络服务需求和服务器配置，IPVS实现了如下八种负载调度算法，这里我们详细讲述最常用的四种调度算法，剩余的四种调度算法请参考其它资料。  
 轮叫调度（Round Robin）  
“轮叫”调度也叫1:1调度，调度器通过“轮叫”调度算法将外部用户请求按顺序1:1的分配到集群中的每个Real Server上，这种算法平等地对待每一台Real Server，而不管服务器上实际的负载状况和连接状态。   
 加权轮叫调度（Weighted Round Robin）   
“加权轮叫”调度算法是根据Real Server的不同处理能力来调度访问请求。可以对每台Real Server设置不同的调度权值，对于性能相对较好的Real Server可以设置较高的权值，而对于处理能力较弱的Real Server，可以设置较低的权值，这样保证了处理能力强的服务器处理更多的访问流量。充分合理的利用了服务器资源。同时，调度器还可以自动查询Real Server的负载情况，并动态地调整其权值。   
 最少链接调度（Least Connections）   
“最少连接”调度算法动态地将网络请求调度到已建立的链接数最少的服务器上。如果集群系统的真实服务器具有相近的系统性能，采用“最小连接”调度算法可以较好地均衡负载。   
 加权最少链接调度（Weighted Least Connections）   
“加权最少链接调度”是“最少连接调度”的超集，每个服务节点可以用相应的权值表示其处理能力，而系统管理员可以动态的设置相应的权值，缺省权值为1，加权最小连接调度在分配新连接请求时尽可能使服务节点的已建立连接数和其权值成正比。  
其它四种调度算法分别为：基于局部性的最少链接（Locality-Based Least Connections）、带复制的基于局部性最少链接（Locality-Based Least Connections with Replication）、目标地址散列（Destination Hashing）和源地址散列（Source Hashing），对于这四种调度算法的含义，本文不再讲述，如果想深入了解这其余四种调度策略的话，可以登陆LVS中文站点zh.linuxvirtualserver.org，查阅更详细的信息。

3.2 高可用性

LVS是一个基于内核级别的应用软件，因此具有很高的处理性能，用LVS构架的负载均衡集群系统具有优秀的处理能力，每个服务节点的故障不会影响整个系统的正常使用，同时又实现负载的合理均衡，使应用具有超高负荷的服务能力，可支持上百万个并发连接请求。如配置百兆网卡，采用VS/TUN或VS/DR调度技术，整个集群系统的吞吐量可高达1Gbits/s；如配置千兆网卡，则系统的最大吞吐量可接近10Gbits/s。

3.3 高可靠性  
LVS负载均衡集群软件已经在企业、学校等行业得到了很好的普及应用，国内外很多大型的、关键性的web站点也都采用了LVS集群软件，所以它的可靠性在实践中得到了很好的证实。有很多以LVS做的负载均衡系统，运行很长时间，从未做过重新启动。这些都说明了LVS的高稳定性和高可靠性。

3.4 适用环境  
LVS对前端Director Server目前仅支持Linux和FreeBSD系统，但是支持大多数的TCP和UDP协议，支持TCP协议的应用有：HTTP，HTTPS ，FTP，SMTP，，POP3，IMAP4，PROXY，LDAP，SSMTP等等。支持UDP协议的应用有：DNS，NTP，ICP，视频、音频流播放协议等。  
LVS对Real Server的操作系统没有任何限制，Real Server可运行在任何支持TCP/IP的操作系统上，包括Linux，各种Unix（如FreeBSD、Sun Solaris、HP Unix等），Mac/OS和Windows等。