

负载均衡高可用的核心概念和常用软件

1、什么是负载均衡高可用

负载均衡建立在现有的网络结构上，他提供了一种廉价、有效、透明的方法来扩大网络设备和服务器的带宽、增加吞吐量、加强网络数据处理能力，以及提高网络的灵活性和可用性。通过负载均衡器，我们可以实现 N 台廉价的 Linux 服务器并行处理，从而达到小型机或大型机的计算能力，这也是为何负载均衡如此流行的主要原因。

高可用（High Availability, HA）其实有两种不同的含义：广义来讲，是指整个系统的高可用性；狭义来讲，一般指主机的冗余接管；

单台负载均衡器位于网站的最前端，他起着分流客户请求的作用，相当于整个网站或系统的人口，如果他出现故障，整个网站也会出现故障，所以这时我们要有一种方案，它能在短时间（一般要求小于 1 秒）内将崩溃的负载均衡器接管过去，这也称为高可用。至于负载均衡器后端的 web 集群、数据库集群，应为了有负载均衡器的内部机制，及时其中的某一台或者两台发生问题，也不会影响整套系统的使用。

我们现在俗称的 Linux 集群，它指的是大范围内的整套系统架构，相当于负载均衡器后的 web 集群、Tomcat 集群或者 MySQL 集群来说，它的涵盖要广得多，包括负载均衡高可用。

目前在线上环境中应用的较多的负载均衡器硬件有 F5 BIG-ip，软件有 LVS NGINX 及 HAproxy，高可用软件有 Heartbeat keepalived，成熟的 Linux 集群架构有 lvs+keepalived nginx | HAProxy+keepalived 及 DRBD+HeartbeatDRBD+

F5 BIG-IP 作为负载均衡器

作为负载均衡器的硬件主要有 F5 BIG-IP 和 Citrix NetScaler，CDN 机房中常见的负载均衡设备就是大名鼎鼎的 F5 BIG-IP。

简单的介绍下，F5 BIG-IP LTM 的官方名称叫做本地流量管理器，可以做 4-7 层的负载均衡，具有负载均衡、应用交换、回话交换、状态监控、智能网络地址转换、通用持续性、响应错误处理、IPV6 网关、高级路由、智能端口镜像、SSL 加速、智能 HTTP 压缩、TCP 优化、第 7 层速率整形、内容缓冲、内容转换、链接加速、告诉缓存、cookie 加密、选择性内容加密、应用攻击过滤、拒绝服务（DOS）攻击和 SYN Flood 保护、包括过滤防火墙、包消毒等功能。

F5 BIG-IP 对 HTTP 负载均衡器的主要功能：

- 通过 12 中灵活的算法将所有流量地分配到各台服务器，对用户而言，他只是一台虚拟服务器。
- 可以确认应用程序能否针对请求返回相应的数据；
- 具有动态 Session 的会话保持功能

LVS 作为负载均衡器

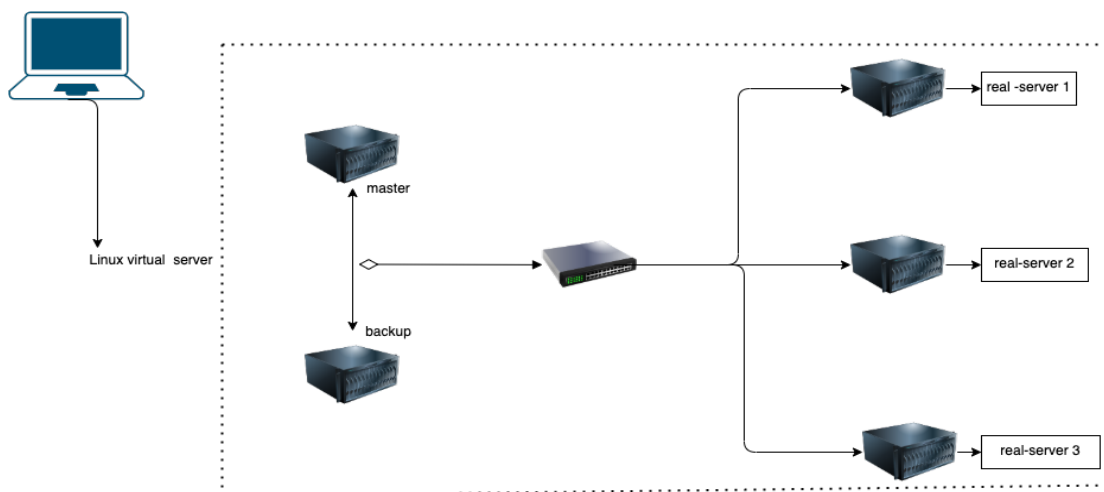
Linux 虚拟服务器是在真实服务器集群上构建的高度可扩展且高度可用的服务器，负载均衡器在 Linux 操作系统上运行。服务器集群的体系结构对最终用户完全透明，并且用户进行交互就像是一台高性能虚拟服务器一样；

LVS 全称为 Linux virtual server，它是章文嵩博士主持的自由软件项目。它建立在一个主控服务器（通常为双击）及若干真实服务器（real-server）所组成的集群之上，real-server 负责提供实际服务。主控端服务器指定的调度算法对 real-server 进行控制。而集群的机构对于用户来说是透明，客户端只于单个的 IP 进行通信，也就是说客户端的视角来看，这里只存在单体服务器。

1) LVS 集群架构

在设计 LVS 集群时需要考虑系统的透明性、可伸缩性、高可用性和易管理性，一般情况下 LVS 集群采用 3 层架构，主要组成部分：

- 负载调度器（load balancer）：它是整个集群对外的前端机，负载将用户的请求发送到一组服务器上执行，而用户认为服务来自一个 IP 地址（我们称之为虚拟 IP 地址）上的；
- 服务器池（server pool）：它是一组真正执行客户请求的服务器，执行的服务有 web 等服务；
- 共享存储（shared storage）：他为服务器池体用一个共享的存储区，这样很容易使得服务器池拥有相同的内容，提供相同的服务；



调度器是服务器集群系统的唯一入口点（single entry point），它可以采用 IP 负载均衡技术和基于内容的请求分发技术，或将两者相结合。在 IP 负载均衡集群技术中，要求服务器池拥有相同的内容，提供相同的服务。

LVS 调度模式

1、NAT 实现虚拟服务器

NAT 的工作原理是报文头（目标地址、源地址和端口等）被正确改写后，客户端相信她们连接的是同一 IP 地址，而不同 IP 地址的服务器组也认为他们是与客户直接相连的。由此，可以用 NAT 方法将不同一个 IP 地址的并行网络服务变成一个在同一 IP 地址上虚拟服务。

NAT 的架构比较简单：在一组服务器前有一个调度器，他们是通过 switch 相连接的，这些服务器提供相同的服务、相同的内容，既不管请求被发送到那一台服务器，执行结构都是一样的。

2、TUN 实现虚拟服务器

基于 NAT 集群中，请求和响应的数据报文都需要通过负载调度器，当真实服务器数量在 10—20 台时，负载调度器将成为整个集群系统的瓶颈。如果能将请求和响应分开处理，即负载调度器只负载调度请求，而响应直接返回给客户端，这将极大提高整个集群系统的吞吐量。

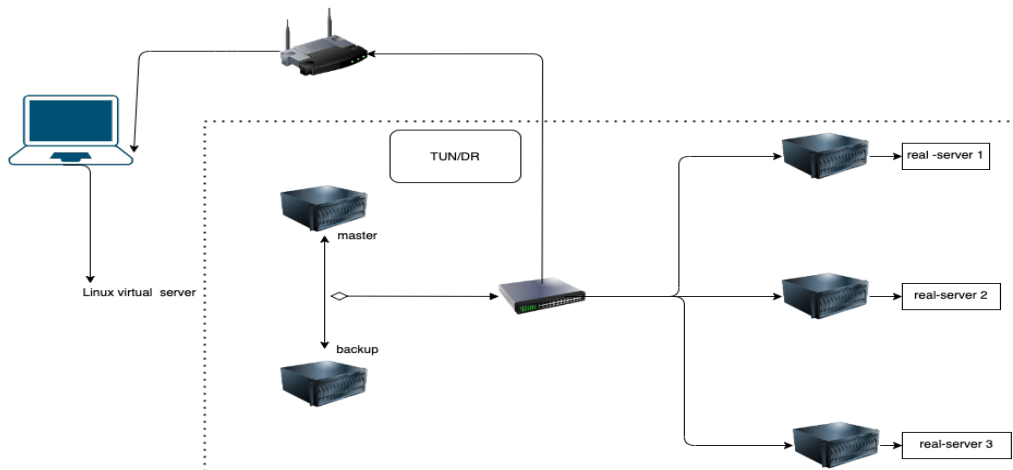
我们利用 IP 隧道技术将请求报文封装并转发给后端服务器，响应报文从后端服务器直接返回给客户端。

IP 隧道（IP tunneling）是将一个 IP 报文封装到另一个 IP 报文中的技术，这可以使目标为一个 IP 地址的数据报文能被封装和转发到另一个 IP 地址。IP 隧道技术也称为 IP 封装技术（IP encapsulation），主要应用于移动主机和虚拟私有网络（virtual private network），在其中隧道都是静态建立的，隧道一端有一个 IP 地址，另一端也有一个唯一的 IP 地址。

3、DR 实现虚拟服务器

DR 于 TUN 方法相同，DR 利用了大多数 Internet 服务的非对称点，负载调度器只负载调度请求，而服务器直接将响应返回给客户端，可以极大地提高整个集群系统吞吐量。

DR 的架构：调度器和服务器组都必须在物理上有一个网卡通过局域网相连，如果同高速的交换机或者 hub 相连。VIP 地址被调度器和服务器组共享，调度器配置的 VIP 地址是对外可见的，用于接收虚拟服务的请求报文。



4、3 种方法的优缺点比较

NAT 的优点是服务器可以运行任何支持 TCP/IP 的操作系统，它只需要一个 IP 地址配置在调度器上，服务器组可以用私有的 IP 地址。缺点是它伸缩能力有限，当服务器节点数目增加到 20 台时，调度器本身有可能成为系统的新瓶颈，因为 NAT 中请求和响应报文都需要通过负载调度器。

TUN 的集群系统中，负载调度器只将请求调度到不同的后端服务器，后端服务器将应答的数据直接返回给用户，这样负载调度器就可以处理大量的请求，它甚至可以调度百台以上的服务器（同等规模的服务器），但它不会成为系统的瓶颈。

DR 调度器只处理客户到服务器端的连接，响应数据可以直接从独立的网络路由放回给客户，这可以极大地提高 LVS 集群系统的可伸缩性。

5、LVS 算法

LVS 提供到 12 种算法

- 1、轮询调度
- 2、加权轮询
- 3、最少连接
- 4、加权最少连接
- 5、基于局部性的最少连接
- 6、带复制的基于局部性的最少连接
- 7、目标地址散列
- 8、源地址散列

Nginx 作为负载均衡服务器

Nginx 即可以作为负载均衡器也可作为反向代理服务器，其配置方法相当简单，可以按轮询、ip_hash、URL_hash、权重等多种方法对后端服务器执行负载均衡操作，同时还支持对后端服务器的健康检查。相对于 LVS 来说，他有一个优势：由于它是基于第七层的负载均衡，是根据包头内的信息来执行负载均衡任务的，所以对网络的依赖性比较小，理论上只要 ping 的通就能够实现负载均衡，Nginx 不仅可以作为一款性能优异的负载均衡器，同时也是一款适用于高并发环境的 web 应用软件；

优点：

- 1) 配置文件非常简单，通俗易懂
- 2) 成本低廉，nginx 是开源软件，可以免费试用；
- 3) 支持 rewrite 重写规则，如果 NGINX Proxy 后端的某台 web 服务器宕机了，不会影响前段访问；
- 4) 节省带宽，支持 GZIP 压缩，可以添加浏览器本地缓存的 Header 头；
- 5) 有内置的健康检查机制；
- 6) 稳定性高，用于反向代理，宕机的概率最小。

HAProxy 负载均衡器

HAProxy 是一款基于 TCP(第四层)和 HTTP（第七层）应用的代理软件，它也可以作为负载均衡器，而且完全免费；借助 HAProxy，可以快速并且可靠的提供基于 TCP 和 HTTP 应用的代理解决方案，HAProxy 最主要的优点是性能突出。

HAProxy 优点：

- 1) 免费且开源；
- 2) 根据官方文档，HAProxy 可以跑满 10Gbps,对于软件级负载均衡而言，这个数字有点大；
- 3) 支持连接拒绝；
- 4) 支持全透明代理，
- 5) HAProxy 现在多用于线上的 MySQL 集群环境，常用它作为 MySQL（读）负载均衡；
- 6) 自带强大的监控服务器状态的页面；
- 7) 支持虚拟主机；