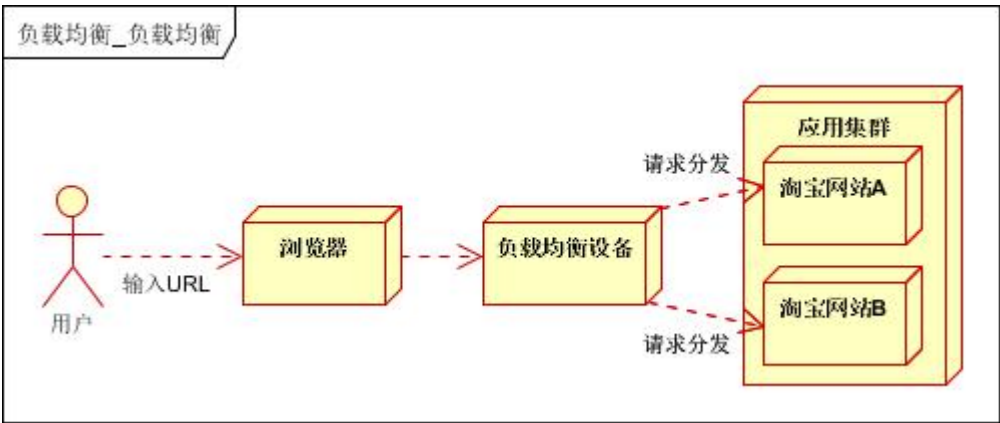


程序拓展常用的负载均衡方法

很多的网站一开始并不需要太大的规模，但是做为网站设计者从一开始就必须考虑到扩展，做一个可扩展性强的架构。

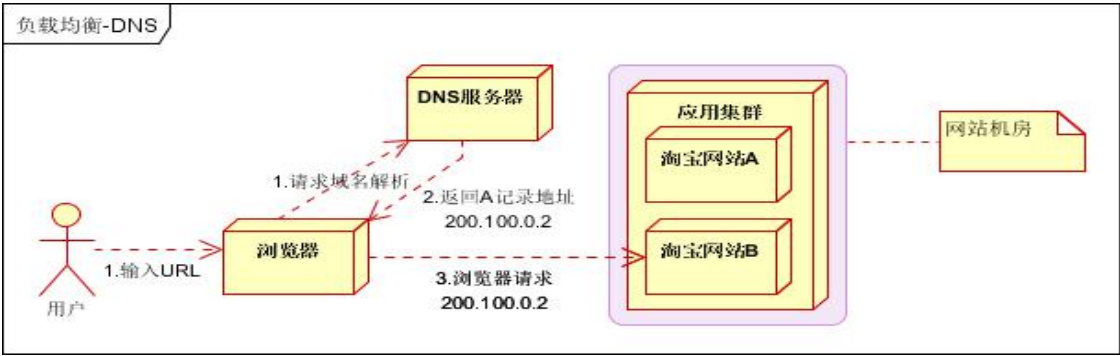
所谓可扩展性就是系统通过规模的扩展来提高系统的承载能力，毕竟服务器自身的垂直扩展很快就会受到制约，单机很快便不能满足我们的需求，因此这种能力往往通过增加物理服务器或集群节点等方面来实现，这种能力越强，承载能力可提升的空间也越大。而 web 站点的水平扩展，负载均衡是一种最常见的手段。



下面介绍一下几种实现负载均衡的方式。

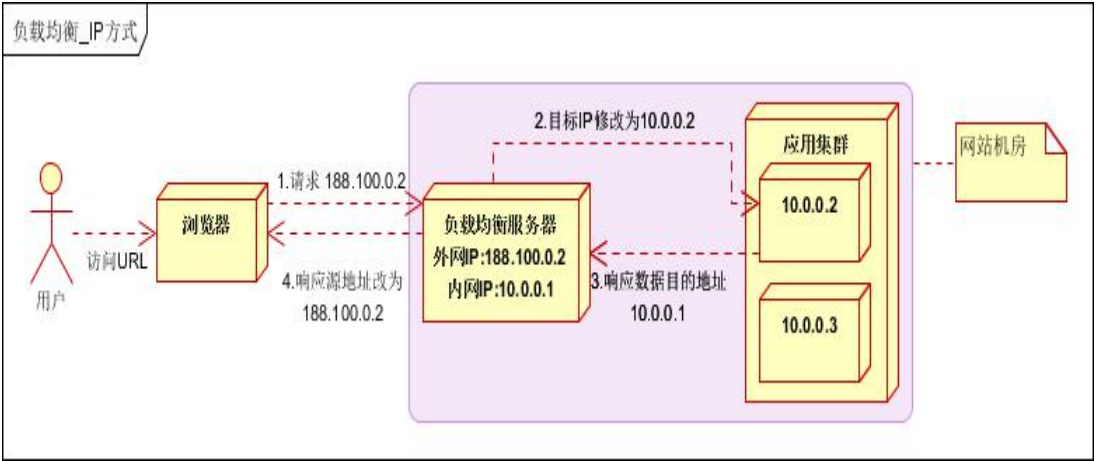
HTTP 重定向 HTTP 重定向，相信对于所有 web 程序员都不陌生，例如我们请求某个页面时，被转向登录页，登录页面之后又被转到了某个页面。大致来说就是，浏览器请求某个 URL 后，服务器通过 HTTP 响应头信息中的 Location 标记返回一个新的 URL，这样浏览器主会继续请求这个新的 URL，完成自动跳转。也正是因为 HTTP 重定向有请求转移和自动跳转的能力，所以我们可以用它来实现负载均衡以实现 WEB 扩展。

DNS 负载均衡 我们知道 DNS 是负责解析域名的，当我们用域名访问站点时，实际上都会经过 DNS 服务器来获取域名指向的 IP，实际上 DNS 服务器完成了域名到 IP 的映射，同样这个映射可以是一对多的，也就是 DNS 可以把对域名的请求按照一定的策略分配到不同的服务器上，这样我们就可以依此来实现负载均衡。貌似和 HTTP 重定向很类似，但是实现的机制却是完全不同的。（在 window 下可以用 nslookup 命令查询域名对应的 IP 地址列表，这个命令会返回离你最近的 DNS 服务器缓存的记录并不一定是全部）：

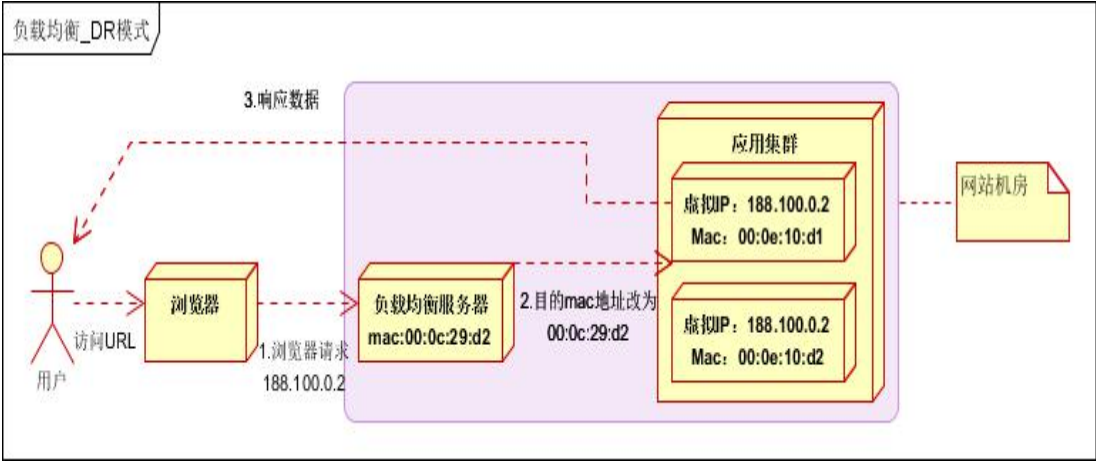


反向代理负载均衡 反向代理服务器，它同样可以作为调度器来实现负载均衡系统。反向代理服务器核心工作是转发 HTTP 请求，它工作在 HTTP 层面，也就是 TCP 七层中的第七层应用层，因此基于反射代理的负载均衡系统也称为七层负载均衡。目前几乎所有主流 WEB 服务器都支持基于反向代理的负载均衡，因此实现它的难度并不大。

IP 负载均衡 基于 IP 负载均衡的系统工作在传输层，会对数据包中的 IP 地址和端口信息进行修改，所以也称为四层负载均衡。它会在数据到达应用层之前，已完成转发，因此这工作都是由系统内核来完成，应用程序对此无能为力，当然性能来说也会很大的提升：



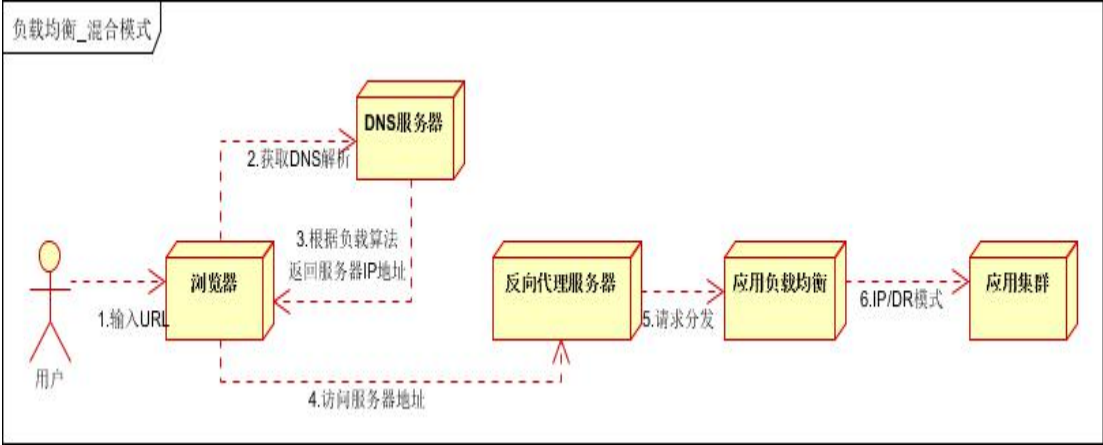
直接路由 不同于 IP 负载均衡，直接路由负载均衡器工作在数据链路层，它通过修改数据包的目标 MAC 地址，将数据包转发到实际服务器，不同的是，这些处理的结果直接发送给用户，不再经过调度器。这时我们的实际服务器必须直接连接到外网，并且不在以调度服务器为默认网关：



IP 隧道 简单的说，就是调度器将收到的数据包封装到一个新的 IP 数据包中，转交给实现服务器，然后实际服务器可以处理数据包直接响应客户端。

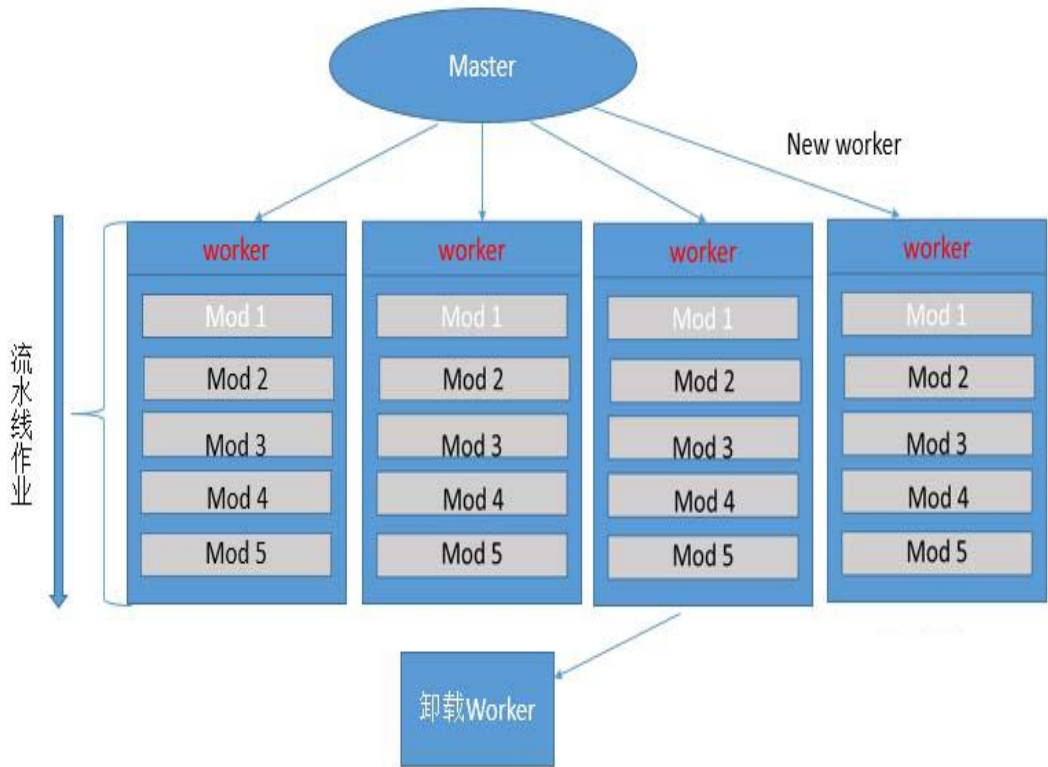
混合型负载均衡 由于多个服务器群内硬件设备、各自的规模、提供的服务等差异，可以考虑给每个服务器群采用最合适的负载均衡方式，然后又在这多个服务器群间再一次负载均衡。

衡或群集起来以一个整体向外界提供服务（即把这多个服务器群当做一个新的服务器群），从而达到最佳的性能。将这种方式称之为混合型负载均衡。

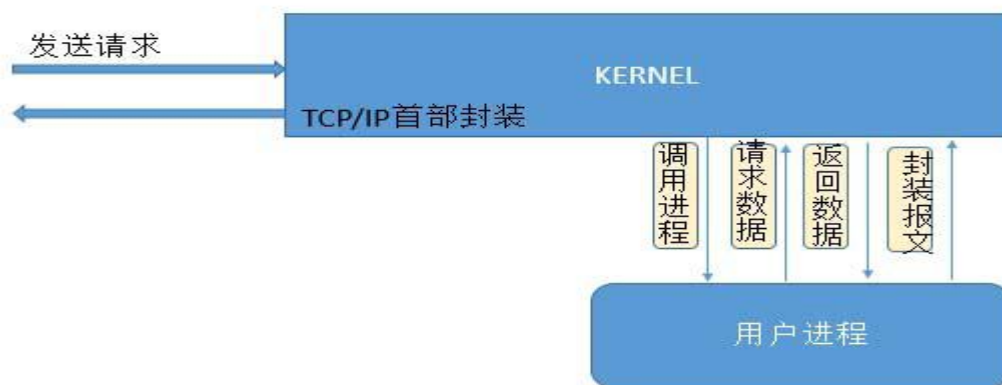


下面介绍业内常用的软件负载均衡软件：Nginx , Lvs , HaProxy 等：

Nginx 是一款轻量级的 Web 服务器/反向代理服务器，工作在七层 Http 协议的负载均衡系统。具有高性能、高并发、低内存使用等特点。是一个轻量级的 Http 和反向代理服务器。Nginx 使用 epoll and kqueue 作为开发模型。能够支持高达 50,000 个并发连接数的响应。它的基本工作模式如下：



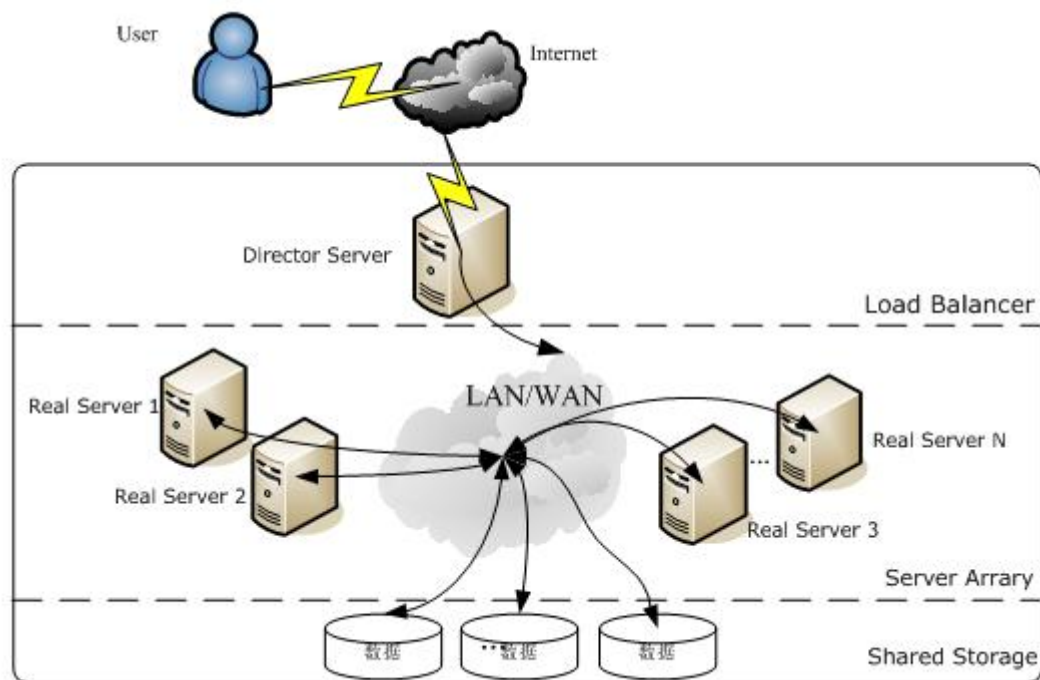
Nginx 支持的 sendfile 机制：



LVS 负载均衡：LVS 是一个开源的软件，由毕业于国防科技大学的章文嵩博士于 1998 年 5 月创立，用来实现 Linux 平台下的简单负载均衡。LVS 是 Linux Virtual Server 的缩写，意思是 Linux 虚拟服务器。

基于 IP 层的负载均衡调度技术，它在操作系统核心层上，将来自 IP 层的 TCP/UDP 请求均衡地转移到不同的服务器，从而将一组服务器构成一个高性能、高可用的虚拟服务器。

它的体系架构图如下：



参考资料：

1. 《常用的负载均衡方法》

<http://www.voidcn.com/blog/wzyzzu/article/p-3484352.html>

2. 大型网站架构系列：负载均衡详解（上）
<http://blog.jobbole.com/97957/>
3. 大型网站架构系列：负载均衡详解（下）
<http://blog.jobbole.com/97960/>