

负载均衡的作用：

1. 解决并发压力，提高应用处理性能（增加吞吐量，加强网络处理能力）；
2. 提供故障转移，实现高可用；
3. 通过添加或减少服务器数量，提供网站伸缩性（扩展性）；
4. 安全防护；（负载均衡设备上做一些过滤，黑白名单等处理）

负载均衡有**硬件**和**软件**两种。

硬件层的比较牛逼,将4-7层负载均衡功能做到一个硬件里面,如F5,梭子鱼,据说yahoo中国!早些时候只用了两台F5做双活.

目前主流的软件负载均衡分为四层和七层

LVS属于四层负载均衡,工作在tcp/ip协议栈上,通过修改网络包的ip地址和端口来转发,由于效率比七层高,一般放在架构的前端.

七层的负载均衡有nginx, haproxy, apache等, 工作在应用层,因此可以将HTTP请求等应用数据发送到具体的应用服务器,如将图片请求转发到特定的服务器上,总之可以做到更智能的负载均衡, 这些功能在四层负载均衡上不好实现,一般放在架构的后面位置,布置在应用服务器前面.

1.1 IP地址重定向

HTTP重定向

Web服务器可通过Http响应头信息中的Location标记来返回一个新的URL，浏览器自动去访问这个新的URL。

实现：可以通过Web应用程序代码实现你想到的调度策略，如可根据请求的URL的不同来进行合理的过滤和转移。

HTTP重定向服务器是一台普通的应用服务器，其唯一的功能就是根据用户的HTTP请求计算一台真实的服务器地址，并将真实的服务器地址写入HTTP重定向响应中(响应状态码302)返回给浏览器，然后浏览器再自动请求真实的服务器。

这种负载均衡方案的优点是比较简单，缺点是浏览器需要每次请求两次服务器才能拿完成一次访问，性能较差;重定向服务器自身的处理能力有可能成为瓶颈。因此这种方案在实际使用中并不见多。

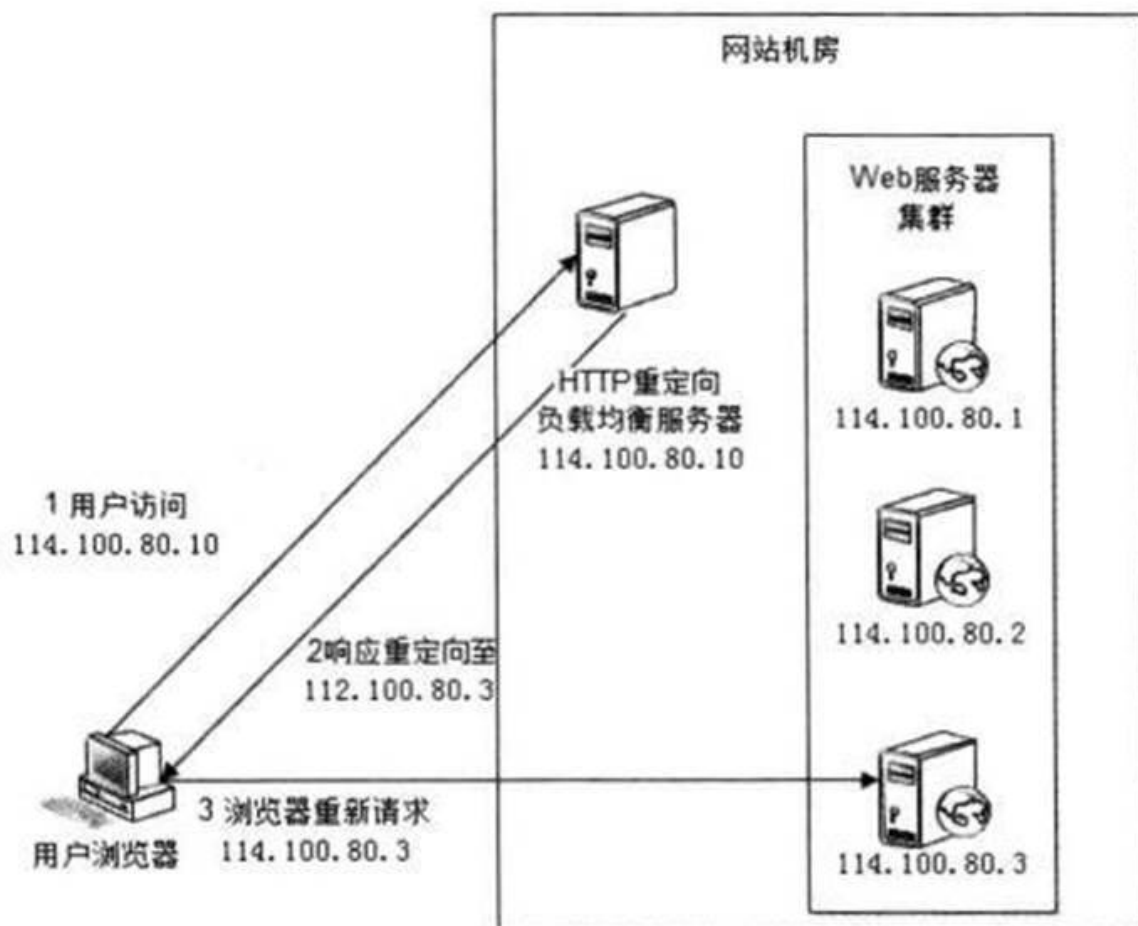


图 6.5 HTTP 重定向负载均衡原理

1.2 DNS负载均衡

DNS负责将用户请求的域名映射为实际的IP地址，这种映射可以是一对多的（DNS的A记录，用来指定域名对应的IP地址），这样DNS服务器便充当负载均衡调度器。大型网站总是部分使用DNS解析，作为第一级负载均衡。

DNS节省了所谓的主站点，DNS服务器已经充当了主站点的职能。常见的策略是对多个A记录进行RR(轮询)

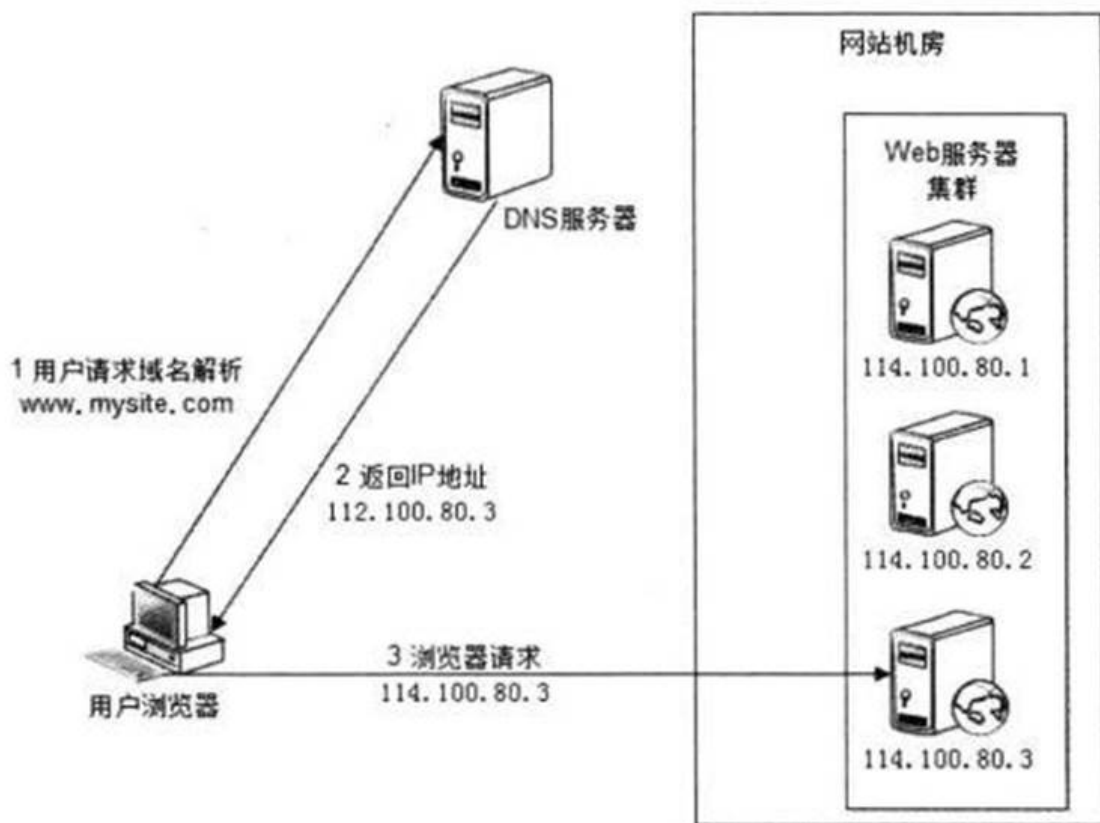


图 6.6 DNS 域名解析负载均衡原理

但在使用DNS负载均衡时，由于DNS数据刷新的延迟问题，无法确保用户请求的完全均衡。而且，一旦其中某台服务器出现故障，即使修改了DNS配置，仍然需要等到新的配置生效后，故障服务器才不会被用户访问到。目前，DNS负载均衡仍在大量使用，但多用于实现“多地就近接入”的应用场景。

优点

1. 使用简单：负载均衡工作，交给DNS服务器处理，省掉了负载均衡服务器维护的麻烦
2. 提高性能：可以支持基于地址的域名解析，解析成距离用户最近的服务器地址，可以加快访问速度，改善性能；

缺点

1. 可用性差：DNS解析是多级解析，新增/修改DNS后，解析时间较长；解析过程中，用户访问网站将失败；
2. 扩展性低：DNS负载均衡的控制权在域名商那里，无法对其做更多的改善和扩展；
3. 维护性差：也不能反映服务器的当前运行状态；支持的算法少；不能区分服务器的差异（不能根据系统与服务的状态来判断负载）

1.3 反向代理服务器

反向代理服务器的核心工作是转发HTTP，它工作在HTTP层面，因此，基于反向代理的负载均衡也称为七层负载均衡。

任何对于实际服务器的HTTP请求都必须经过调度器；调度器必须等待实际服务器的HTTP响应，并将它反馈给用户。

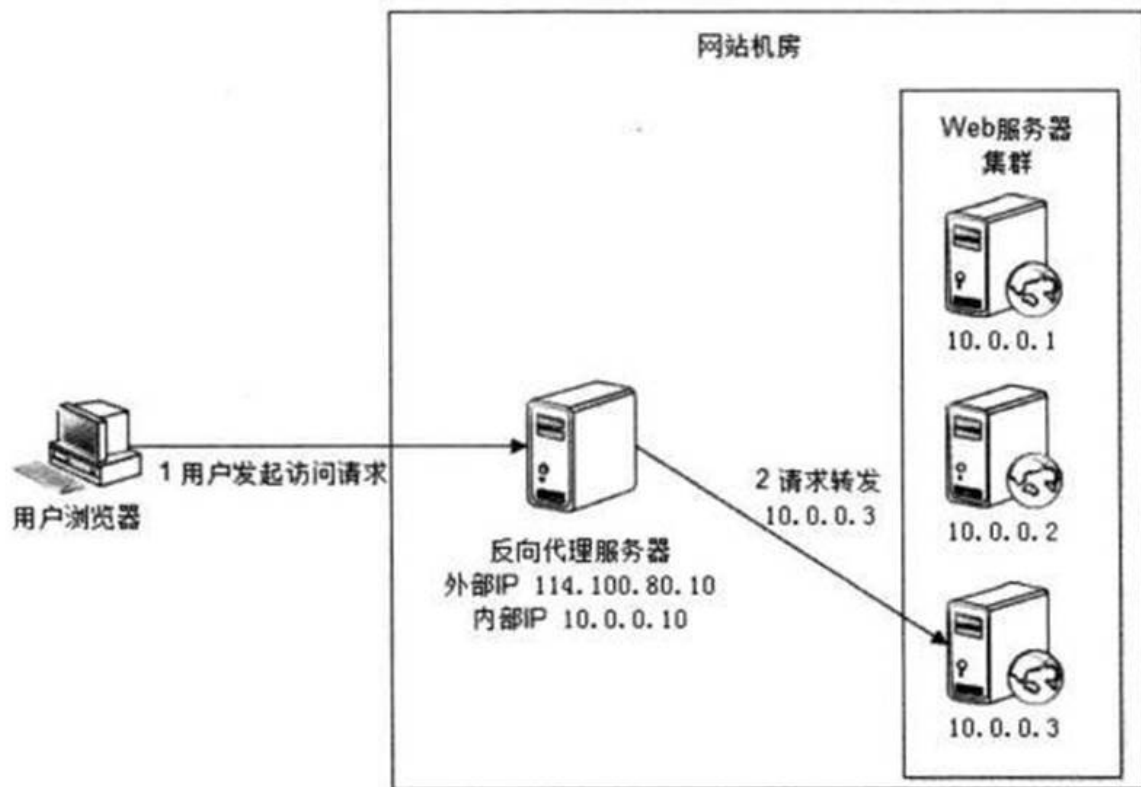


图 6.7 反向代理负载均衡原理

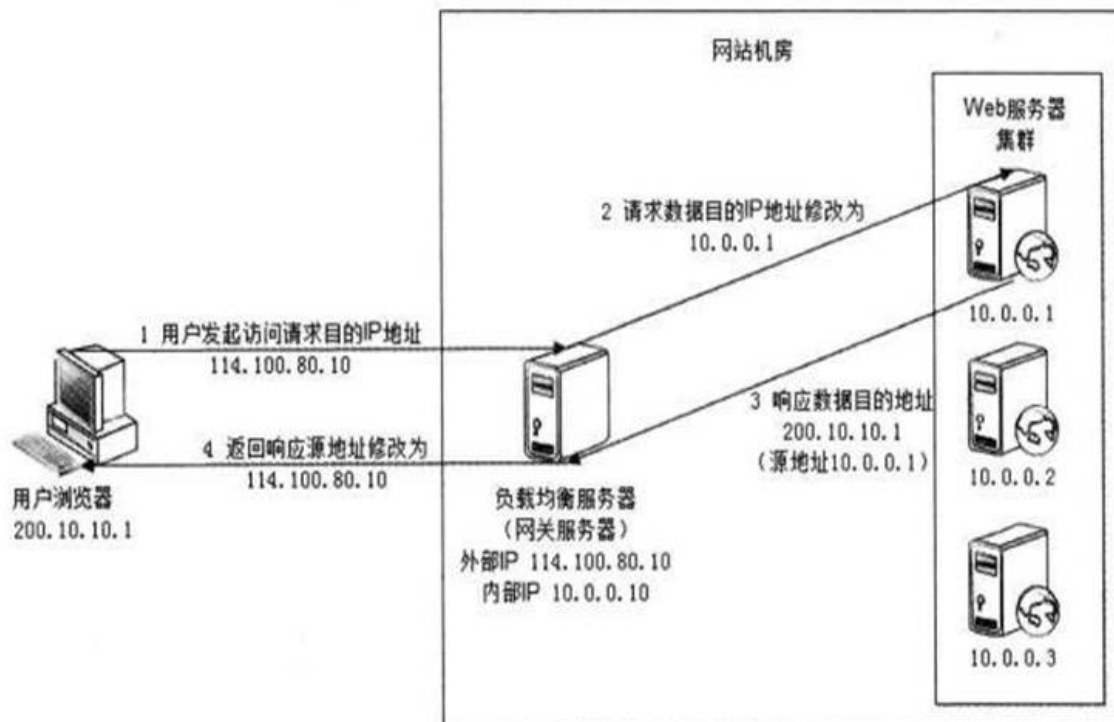
优点：部署简单，处于http协议层面。

缺点：使用了反向代理服务器后，web 服务器地址不能直接暴露在外，因此web服务器不需要使用外部IP地址，而反向代理服务作为沟通桥梁就需要配置双网卡、外部内部两套IP地址。不适合视频类，直播类等高流量业务

1.4 IP负载均衡

网络地址转换(NAT)负载均衡工作在传输层，对数据包中的IP地址和端口进行修改，从而达到转发的目的，称为四层负载均衡。

NAT服务器（前端服务器）必须作为实际服务器（后端服务器）的网关，否则数据包被转发后将一去不返。



优点:

在内核进程完成数据分发，比在应用层分发性能更好。

缺点:

所有请求响应都需要经过负载均衡服务器，集群最大吞吐量受限于负载均衡服务器网卡带宽。

1.5 直接路由（数据链路层）

这种方式工作在数据链路层。它修改数据包的目标MAC地址，并没有修改目标IP（因为这种转发工作在数据链路层，它对上层端口无能为力），然后发给实际的服务器，实际服务器的响应数据直接发回给用户，而不用经过调度器。但实际服务器必须接入外网，而且不能将调度器作为默认网关，要给实际服务器添加和调度器IP地址相同的IP别名。

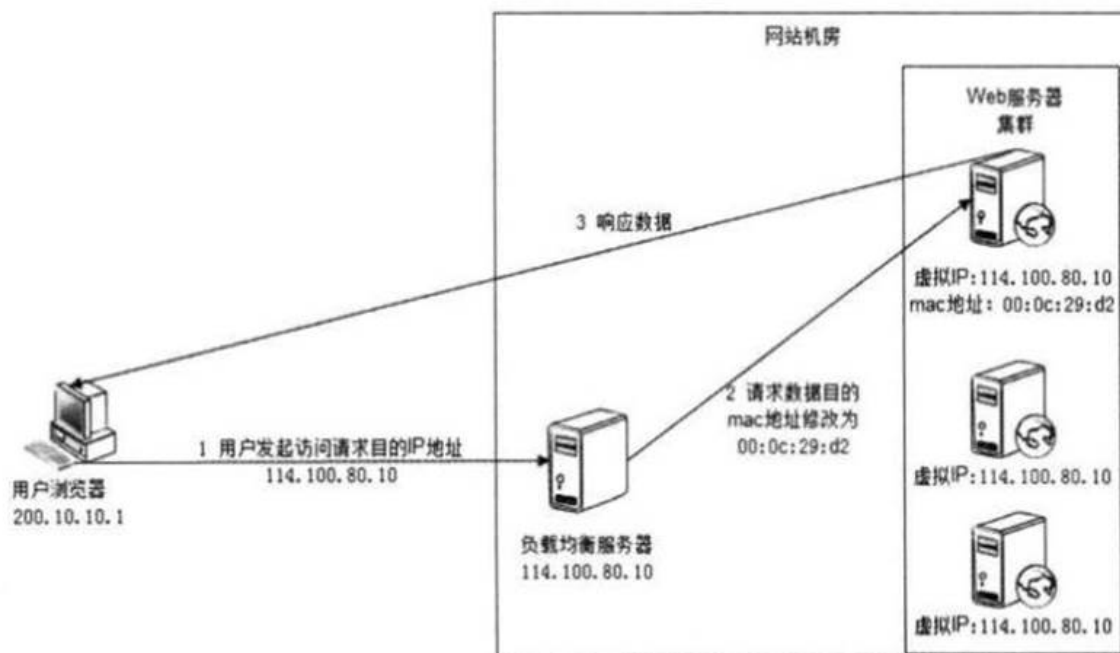


图 6.9 数据链路层负载均衡原理

优点：不需要负载均衡服务器进行地址的转换。数据响应时不需要经过负载均衡服务器。

缺点：配置复杂。

1.6 IP隧道

基于IP隧道的负载均衡系统也可以使用LVS（Linux Virtual Server）来实现，称为LVS-TUN。与LVS-DR不同的是，实际服务器和调度器可以不在同一个WAN网段，调度器通过IP隧道技术来转发请求到实际服务器，所以实际服务器必须有合法的IP地址。

基于IP隧道的请求转发机制，是将调度器收到的IP数据包封装在一个新的IP数据包中，转交给实际服务器，然后实际服务器的响应数据包可以直接到达用户端。

基于IP隧道的独特方式，可以将实际服务器部署在不同的地域并根据就近原则转移请求，比如一些CDN服务器就是基于IP隧道技术实现的。

单网卡10G转发技术

<https://blog.csdn.net/chao199512/article/details/81293490>

<https://zhuanlan.zhihu.com/p/22360384>