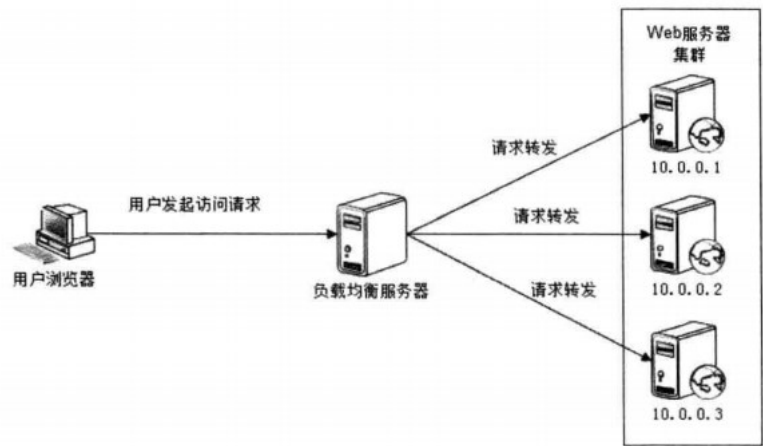


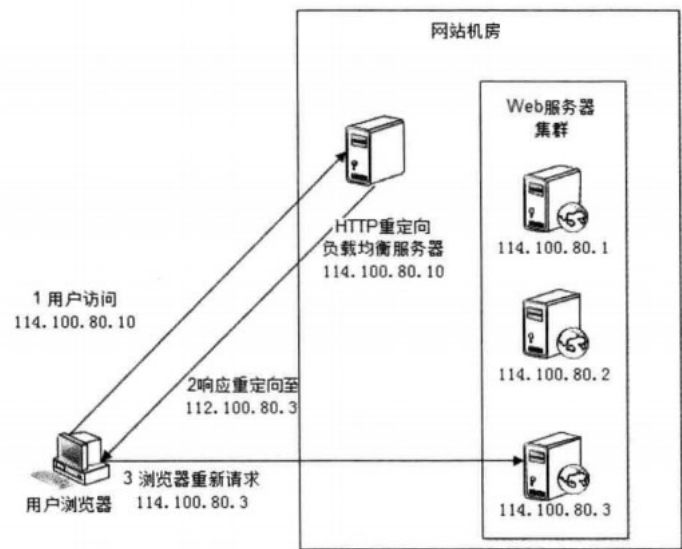
负载均衡的基础技术种类

负载均衡的基础技术种类



负载均衡实现应用服务器伸缩性

- 1、 HTTP重定向负载均衡
- 利用HTTP重定向协议实现负载均衡。如下图所示：



HTTP重定向负载均衡原理

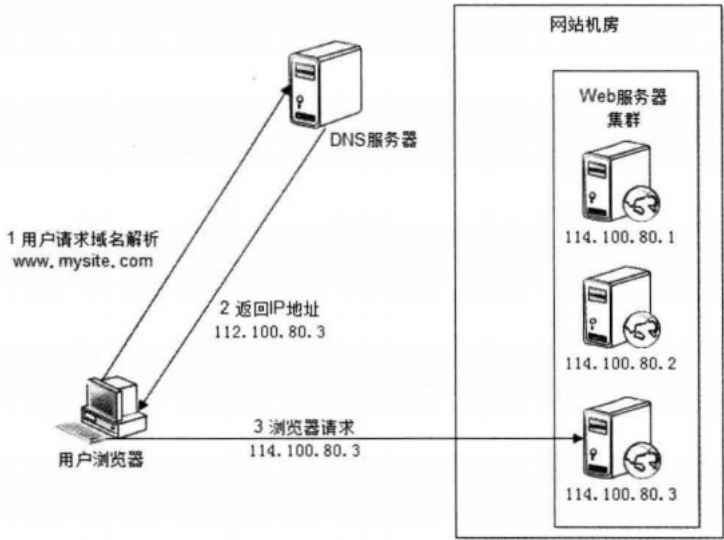
HTTP重定向服务器是一台普通的应用服务器，其唯一的功能就是根据用户的HTTP请求计算一台真实的Web服务器地址，并将该Web服务器地址写入HTTP重定向响应中(响应状态码302)返回给用户浏览器。

缺点：

- 1、浏览器需要两次请求服务器才能完成一次访问，性能较差
- 2、重定向服务器自身的处理能力有可能成为瓶颈，整个集群的伸缩性规模有限
- 3、使用HTTP302响应码重定向，有可能使搜索引擎判断为SEO作弊，降低搜索排名

2、DNS域名解析负载均衡

这是利用DNS处理域名解析请求的同时进行负载均衡处理的一种方案，如下图所示：



DNS域名解析负载均衡原理

在DNS服务器中配置多个A记录，如：[www.mysite.com](#) IN A 114.100.80.1、[www.mysite.com](#) IN A 114.100.80.2、[www.mysite.com](#) IN A 114.100.80.3。

每次域名解析请求都会根据负载均衡算法计算一个不同的IP地址返回，这样A记录中配置的多个服务器就构成一个集群，并可以实现负载均衡。上图中的浏览器请求解析域名[www.mysite.com](#)，DNS根据A记录和负载均衡算法计算得到一个IP地址114.100.80.3，并返回给浏览器；浏览器根据该IP地址，访问真是物理服务器114.100.80.3。

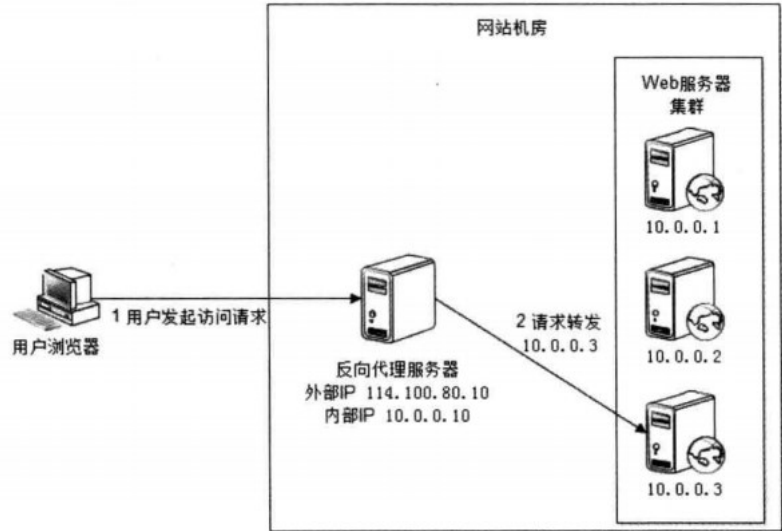
优点：将负载均衡的工作转交给DNS，省掉了网站管理维护负载均衡服务器的麻烦，同时许多DNS还支持基于地理位置的域名解析，即将域名解析成距离用户地理位置最近的一个服务器地址，这样可加快用户访问速度，改善性能。

缺点：目前的DNS是多级解析，每一级DNS都可能缓存A记录，当下线某台服务器后，即使修改了DNS的A记录，要使其生效也需要较长的时间，这段时间，DNS依然会将域名解析到已经下线的服务器，导致用户访问失败；而且DNS负载均衡的控制权在域名服务商那里，网站无法对其做更多改善和更强大的管理。

大型网站总是部分使用DNS域名解析，利用域名解析作为第一级负载均衡手段，即域名解析得到的一组服务器并不是实际提供Web服务的物理服务器，而是同样提供负载均衡的内部服务器，这组内部负载均衡服务器再进行负载均衡，将请求分发到真实的Web服务器上。

3、反向代理负载均衡（由于反向代理服务器转发请求在HTTP协议层面，因此也叫应用层负载均衡）

利用反向代理服务器进行负载均衡，如下图：



反向代理负载均衡原理

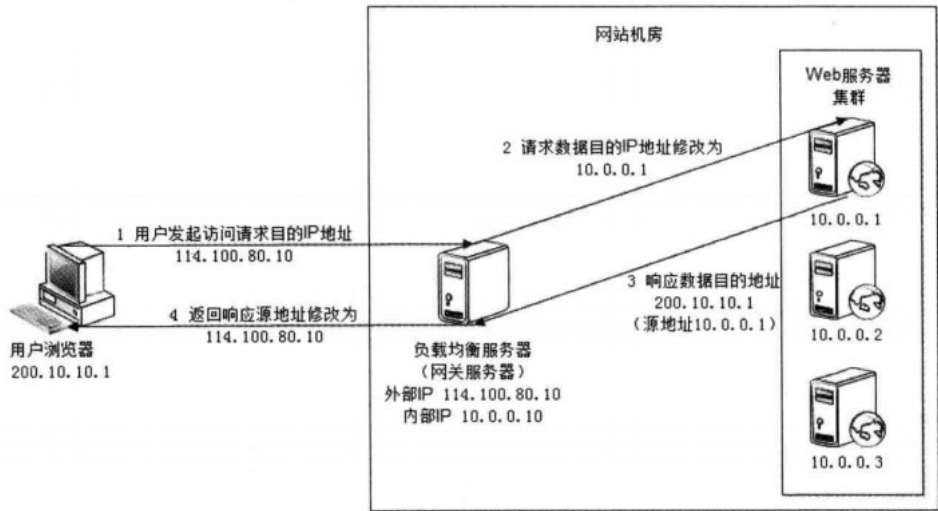
前面我们提到利用反向代理缓存资源，以改善网站的性能。实际上，在部署位置上，反向代理服务器处于Web服务器前面（这样才能缓存Web响应，加速访问），这个位置正好是负载均衡服务器的位置，所以大多数反向代理服务器同时提供负载均衡的功能，管理一组Web服务器，将请求根据负载均衡算法转发到不同的Web服务器上。Web服务器处理完成的响应也需要通过反向代理服务器返回给用户。由于Web服务器不直接对外提供访问，因此Web服务器不需要使用外部IP地址，而反向代理服务器则需要配置双网卡和内部外部两套IP地址。

上图中，浏览器访问清酒的地址是反省代理服务器的地址114.100.80.10，反向代理服务器收到请求后，根据负载均衡算法计算得到一台真实物理服务器的地址10.0.0.3，并将请求转发给服务器。10.0.0.3处理完请求后将响应返回给反向代理服务器，反向代理服务器再将该响应返回给用户。

- 优点：和反向代理服务器功能集成在一起，部署简单
- 缺点：反向代理服务器是所有请求和响应的中转站，其性能可能会成为瓶颈

4、IP负载均衡

在网络层通过修改请求目标地址进行负载均衡，如下图所示：



IP负载均衡原理

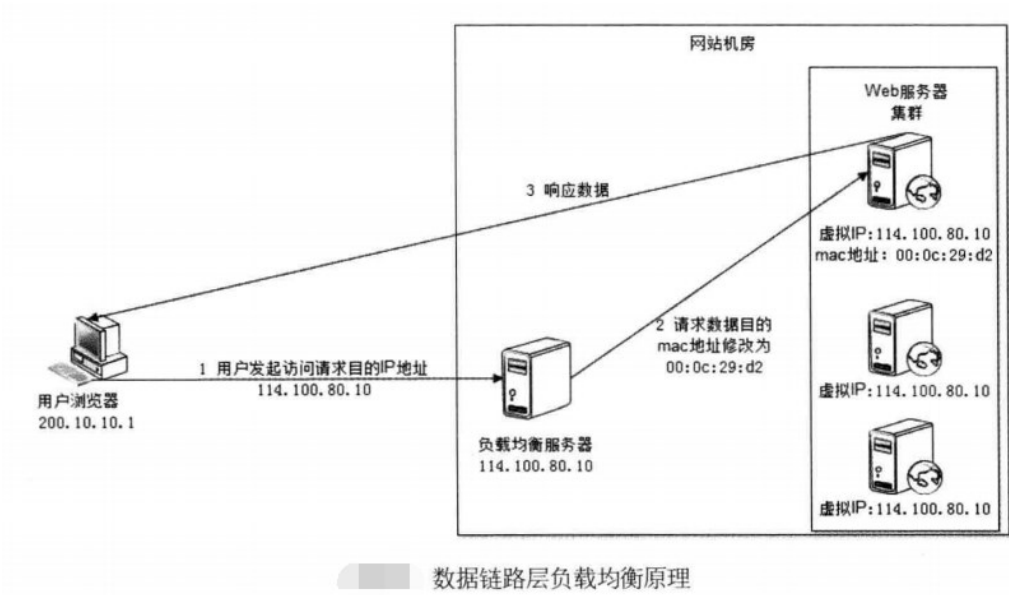
真实物理Web服务器响应数据包返回给 负载均衡服务器方案：

- 1：负载均衡服务器在修改目的的IP地址的同时修改源地址，将数据源地址设为自身IP，即源地址转换(SNAT),这样Web服务器的响应会再回到负载均衡服务器
- 2：将负载均衡服务器同时作为物理服务器集群的网关服务器，这样所有响应数据都会到达负载均衡服务器

IP负载均衡在内核进程完成数据分发，较反向代理 负载均衡(在应用程序中分发数据)有更好的处理能力。但是由于所有请求响应都需要经过负载均衡服务器，集群的最大响应数据吞吐量不得受制于负载均衡服务器 网卡带宽。对于提供下载服务或者视频服务等需要传输大量 数据的网站而言，难以满足需求。

5、数据链路层负载均衡

顾名思义，数据链路层负载均衡是指在通信协议的数据 链路层修改mac地址进行负载均衡，如下图所示：



6、负载均衡算法

负载均衡服务器的实现可以分为两个部分：

- 1：根据负载均衡算法和Web服务器列表计算得到集群中一台Web服务器的地址
- 2：将请求数据发送到该地址对应的Web服务器上

前面描述了如何将请求数据发送到Web服务器，二具体的负载均衡算法通常有以下几种

轮询（Round Robin，RR）

所有请求被一次分发到每台应用服务器上，即每台服务器需要处理的请求数目都相同，适合于素有服务器硬件都相同的场景。

加权轮询（Weighted Round Robin，WRR）

根据应用服务器硬件性能的情况，在 轮询的基础上，按照配置的权重将请求分发到每个服务器，高性能的服务器能分配更多请求。

随机（Random）

请求被随机分配到各个应用服务器，在许多场合下，这种方案都很简单实用，因为好的随机数本身就很均衡。即使应用服务器硬件配置不同，也可以使用加权随机算法。

最少连接（Least Connections）

记录每个应用服务器正在处理额连接数（请求数），将新到的请求分发到最少连接的服务器上，应该说，这是最符合负载均衡定义的算法。同样，最少连接算法也可以实现加权最少连接。

源地址散列（Source hashing）

根据请求来源的IP地址进行Hash计算，得到应用服务器，这样来自同一个IP地址的请求总在同一个服务器上处理，该请求的上下文信息可以存储在这台服务器上，在一个会话周期内重复使用，从而实现会话黏滞。

标签: 负载均衡, 网站架构

好文要顶

关注我

收藏该文



zero.huang

关注 - 8

粉丝 - 5

2

0

+加关注

« 上一篇: [建设你的知识结构: 一份阅读书目单](#)

posted @ 2016-03-29 16:39 zero.huang 阅读(1583) 评论(1) 编辑 收藏

评论

#1楼 2017-03-25 20:17 | xiaohuan1

博主好，我对这个内容有点不太懂：
4.IP负载均衡，真实物理Web服务器响应数据包返回给 负载均衡服务器方案：
这里有两个方案，不管是哪个方案，web服务器响应数据包，都会返回到负载均衡服务器中。那么这两个方案有什么区别呢
支持(0) 反对(0)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】超50万C++/C#源码: 大型实时仿真HMI组态CAD\GIS图形源码！
【推荐】专业便捷的企业级代码托管服务 - Gitee 码云

相关博文：

- [负载均衡技术（一）——负载均衡技术介绍](#)
- [负载均衡的几种类型](#)
- [群集的负载均衡技术](#)
- [\[负载均衡\] 简述负载均衡&CDN技术](#)
- [负载均衡技术介绍](#)

最新新闻：

- [永不造车？拆解华为拥有的自动驾驶和电动汽车关键技术](#)
 - [6天面试、斩获6家硅谷巨头Offer，我是如何做到的？](#)
 - [IBM成Z代人最青睐科技公司 谷歌和亚马逊分列第二三名](#)
 - [为什么我们更像是在为抖音筛选内容，而非消费内容？](#)
 - [人人车变脸：曾经想干掉黄牛，如今成了“黄牛公司”](#)
- » [更多新闻...](#)