

昵称： 大鹏展翅seu
园龄： 4年6个月
粉丝： 0
关注： 0
[+加关注](#)

2019年2月						
日	一	二	三	四	五	六
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	1	2
3	4	5	6	7	8	9

搜索

找找看

谷歌搜索

常用链接

- [我的随笔](#)
- [我的评论](#)
- [我的参与](#)
- [最新评论](#)
- [我的标签](#)

随笔档案

- [2014年11月 \(1\)](#)
- [2014年10月 \(1\)](#)
- [2014年9月 \(2\)](#)
- [2014年8月 \(1\)](#)

阅读排行榜

- [1. HTTP服务负载均衡总结\(268\)](#)
- [2. HTTP缓存机制梳理\(87\)](#)
- [3. Hello world\(47\)](#)
- [4. 归并排序\(30\)](#)
- [5. 堆排序\(21\)](#)

HTTP服务负载均衡总结

从一开始就要思考扩展的架构，所谓可扩展性指的是通过扩展规模提高承载能力的本领，往往体现在增加物理服务器或者集群节点。负载均衡是常见的水平扩展的手段。

目标：（1）减少单点故障（2）提升整体吞吐量（3）根据负载分配任务量

HTTP重定向

通过HTTP Location重定向，通过一定的策略（如地域就近）将主站URL转移到分站/镜像URL，分散主站工作量。实现负载均衡的同时，加快用户下载速度，适用于网络IO成为瓶颈时（如下载服务）。

HTTP重定向通常采用随机策略。如果使用轮询（RR），需要记录最后一次重定向的服务器序号，给通常无状态的HTTP带来额外开销；由于需要锁机制保证任何时刻只有一个请求能修改它，对并发性也有一定影响。

根据概率统计理论，随着吞吐率的增加，随机调度也会逐渐趋近于顺序调度的均衡效果。

多数情况，重定向向负载均衡不好控制，无法阻止用户直接访问真实服务器。只有类似文件下载、广告展示等一次性的请求适合重定向。

DNS

DNS提供域名解析服务，即域名到ip的映射，这种映射可以是一对多的，这时DNS便起了负载均衡的作用。DNS服务器软件提供了丰富的调度策略，最常用的是根据用户IP进行智能解析，从所有可用的A记录中找到离用户最近的一台服务器。利用动态DNS，可以实现故障转移，但因为DNS的层级缓存，会存在一定延迟。

除了因为缓存带来的更新延迟，由于负载均衡工作在DNS层面，调度灵活性较差，无法将HTTP请求的上下文引入调度策略。

反向代理

反向代理服务器的核心工作是转发HTTP请求，所以又叫七层负载均衡。相比HTTP重定向和DNS，反向代理是**转发**请求，而前者是**转移**。所有请求都经过反向代理，等待实际服务器（也叫后端服务器）的响应，再反馈给用户。这样可以将调度策略落实到每一个HTTP请求，从而实现更加可控的负载均衡策略。

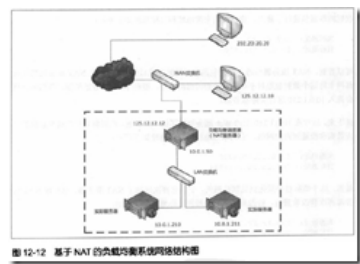
反向代理的扩展能力，一方面是自身并发处理的能力，当反向代理服务器吞吐率接近极限时，无论添加多少后端服务器也无济于事。另一方面，转发操作本身需要一定开销，如创建线程、与后端服务器建立TCP连接、接收后端返回的处理结果、分析HTTP头信息、用户空间和内核空间的频繁切换等，通常这部分时间不长，但当后端服务器处理请求时间非常短时（如静态资源），转发的开销就显现出来。

反向代理一般都有监控后端服务器的功能，可以简单配置健康探测，调度器会自动放弃异常的后端服务器。实际应用可以部署一定数量的备用服务器，当一些后端服务器异常时接替它们的工作，保证整体性能。

如果后端服务器是有状态的，如在session中保存了用户数据，需要使用粘滞会话（Sticky Sessions）。一种方法是使用用户的IP地址作为标识，将IP地址做散列计算映射到后端服务器上。还可以利用Cookie，反向代理将后端服务器的编号追加写到用户的Cookie中，这样便知道下次请求该发给哪个后端服务器。粘滞会话可能会破坏负载均衡策略，使得后端服务器个性化，尽量避免在后端服务器存储状态，采用分布式Session或分布式缓存来替代。

IP

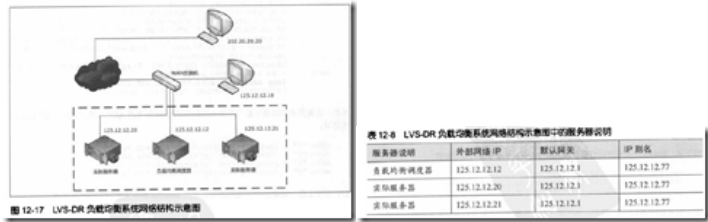
反向代理本身的开销制约了其扩展性，考虑将调度器工作在应用层以下，网络数据包从内核缓冲区进入用户地址空间之前，Linux内核就将其转发到实际服务器上。



NAT工作在传输层，可以修改数据包的IP地址和端口，所以也叫四层负载均衡。NAT服务器负责连接外部网络和内部网络，实际服务器必须将NAT服务器作为默认网关。当数据包到达NAT服务器，NAT服务器将目标地址从NAT的地址修改为实际服务器地址的IP和端口，然后指定内部网卡将数据包投递到内部网络；实际服务器收到数据包进行处理，数据包到达默认网关，即NAT服务器，NAT服务器再修改数据包，将源地址从实际服务器地址改为NAT服务器。Linux使用NetFilter/iptables维护数据包路由规则。使用IPVS（LVS）配置基于NAT的负载均衡。由于在内核转发的低开销，吞吐率比反向代理高，转发能力主要取决于NAT服务器的网络带宽。



对于下载或视频等站点，NAT服务器的带宽会成为瓶颈。LVS提供了直接路由（Direct Router）的方式，它的负载均衡工作在数据链路层，通过修改数据包的MAC地址，将数据包转发到实际服务器，而且响应数据包直接发给用户端，不用经过调度器，所以实际服务器必须直接接入外部网络，也不能将调度器设为默认网关。直接路由使用IP别名，需要给实际服务器添加和调度器IP地址相同的IP别名，这样可以正常将请求数据包转发到实际服务器。通常响应数据包比请求数据包大的多，所以LVS-DR的优势就是响应数据包不用经过调度器，避免了调度器的带宽瓶颈。



好文要顶 关注我 收藏该文





大鹏展翅seu

关注 - 0

粉丝 - 0

+加关注

« 上一篇: [HTTP缓存机制梳理](#)

posted @ 2014-11-12 11:20 大鹏展翅seu 阅读(268) 评论(0) 编辑 收藏

[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：

- [LVS负载均衡服务](#)
- [负载均衡与HTTP加速](#)
- [负载均衡技术\(二\) ——常用负载均衡服务介绍](#)
- [nginx负载均衡总结](#)
- [负载均衡之HTTP重定向](#)

最新新闻：

- [张一鸣豪赌千亿营收，但字节跳动仍将面临三重难关](#)
 - [马斯克私有化推文影响犹在 特斯拉还在应付股东集体诉讼案](#)
 - [字节跳动新增商标“字节锤子”](#)
 - [记者探访FF美国总部：贾跃亭没忘记乐视债务，也不愿放弃造车梦](#)
 - [MIT和微软的新成果，能否帮自动驾驶摆脱成长的烦恼？](#)
- » [更多新闻...](#)

