

系统架构之负载均衡

深耕IT

百家号 | 17-04-29 21:41

在做系统架构规划的时候，负载均衡，HA(高可用性集群，是保证业务连续性的有效解决方案，一般有两个或两个以上的节点，且分为活动节点及备用节点，当活动节点出现故障的时候，由备用节点接管)都是经常需要考虑的方案。对应并发及单点故障，考虑负载均衡方案是必不可少的。如果并发不高只是应对单点故障，则通常使用HA方案。

负载均衡(LoadBalance)是集群技术(Cluster)的一种应用。负载均衡可以将工作任务分摊到多个处理单元，从而提高并发处理能力。目前最常见的负载均衡应用是Web负载均衡。根据实现的原理不同，常见的web负载均衡技术包括：DNS轮询、IP负载均衡和CDN。其中IP负载均衡可以使用硬件设备或软件方式来实现。

(一)、硬件方式

采用F5服务器来做负载均衡，F5的全称是F5-BIG-IP-GTM，是最流行的硬件负载均衡设备，其并发能力达到百万级。F5的主要特性包括：

多链路的负载均衡和冗余

可以接入多条ISP链路，在链路之间实现负载均衡和高可用。

防火墙负载均衡

F5具有异构防火墙的负载均衡与故障自动排除能力。

服务器负载均衡

这是F5最主要的功能，F5可以配置针对所有的对外提供服务的服务器配置VirtualServer实现负载均衡、健康检查、回话保持等。

高可用

F5设备自身的冗余设计能够保证99.999%的正常运行时间，双机F5的故障切换时间为毫秒级。

使用F5可以配置整个集群的链路冗余和服务器冗余，提高可靠的健康检查机制，以保证高可用。

安全性

与防火墙类似，F5采用缺省拒绝策略，可以为任何站点增加额外的安全保护，防御普通网络攻击，包括DDoS、IP欺骗、SYN攻击、teartop和land攻击、ICMP攻击等。

易于管理

深耕IT

百家号 | 最近更新:17-04-29 21:41

简介:科技改变世界，IT支撑科技

作者最新文章

金融业务管理系统架构设计之贷后检查方案


金融业务管理系统架构设计之产品管理方案

金融业务管理系统架构设计之文档处理方案


相关文章




2019年云计算十大趋势：云成为获取人工...
雷锋网 02-21




跑男团新老成员首聚，杨颖跟宋雨琦斗舞，...
手机中国网 02-21



坂口健太郎与趣里因合作日剧迅速熟悉 近彻...
手机中国网 02-21



北京电影学院撤销翟天临博士学位 论文抄袭...
北京日报客户端 02-20



昆明今年在建高速公路达12条 长龙高速拟年...
光明网 02-21

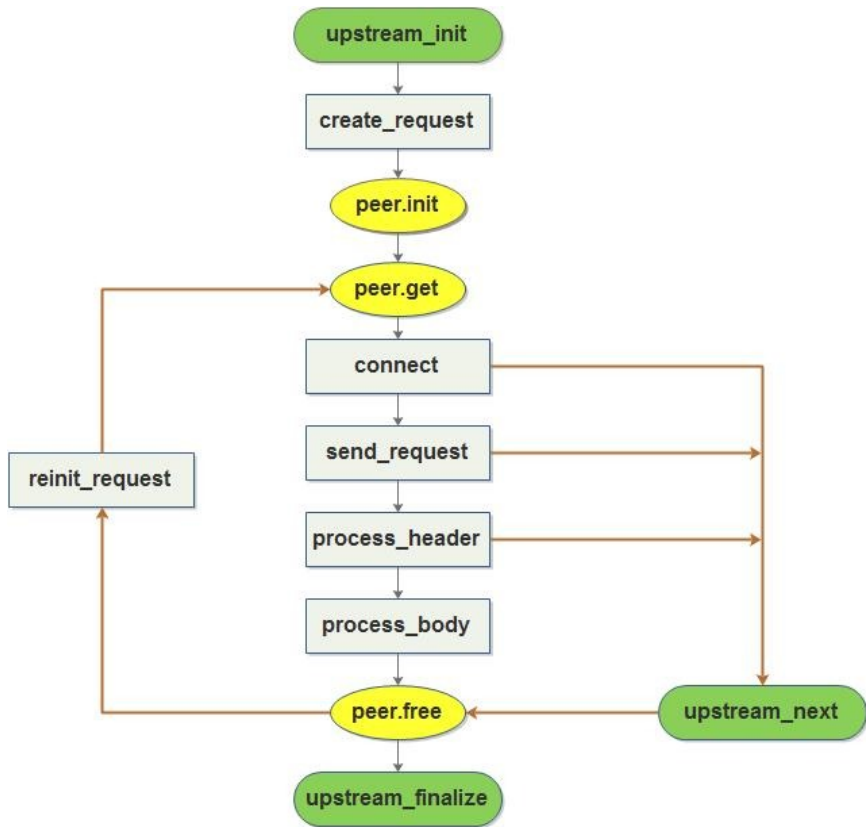
其他

F5还提供了SSL加速、软件升级、IP地址过滤、带宽控制等辅助功能。

(二)、软件方式

(1) nginx

nginx我们常用于做反向代理，但它的负载均衡能力也很强。



nginx负载均衡流程图

Nginx目前提供的负载均衡模块：

ngx_http_upstream_round_robin，加权轮询，可均分请求，是默认的HTTP负载均衡算法，集成在框架中。

ngx_http_upstream_ip_hash_module，IP哈希，可保持会话。

ngx_http_upstream_least_conn_module，最少连接数，可均分连接。

ngx_http_upstream_hash_module，一致性哈希，可减少缓存数据的失效。

(2) apachehttpserver

apachehttpserver作为LoadBalance前置机分别有三种不同的部署方式，分别是：

1、轮询均衡策略的配置

(3) LVS

LVS是LinuxVirtualServer的简称，也就是Linux虚拟服务器,是一个由章文嵩博士发起的自由软件项目，它的官方站点是www.linuxvirtualserver.org。现在LVS已经是Linux标准内核的一部分，在Linux2.4内核以前，使用LVS时必须重新编译内核以支持LVS功能模块，但是从Linux2.4内核以后，已经完全内置了LVS的各个功能模块，无需给内核打任何补丁，可以直接使用LVS提供的各种功能。LVS支持多种负载均衡机制。包括：VS/NAT（基于网络地址转换技术）、VS/TUN（基于IP隧道技术）和VS/DR（基于直接路由技术）。此外，为了适应不同的需要，淘宝开发了VS/FULLNAT，从本质上来说也是基于网络地址转换技术。最近还有一个基于VS/FULLNAT的DNAT模块。不管使用哪种机制，LVS都不直接处理请求，而是将请求转发到后面真正的服务器(RealServer)。不同的机制，决定了响应包如何返回到客户端。

LVS调度算法：

轮叫调度

调度器通过外部请求的顺序轮流的分配到集群中的真实服务器上，对每台服务器都是均等的。但是这样调度器不会考虑服务器上实际的连接数和系统负载，导致服务器处理请求慢，系统负载增大。

加权轮叫

调度器通过一个算法根据真实服务器的不同处理能力来分配访问请求，这样可以保证服务器的处理能力。

最少连接

调度器将访问请求自动的分配到已建立连接最少的服务器上，如果在集群中每台服务器的性能差不多的话，则这种算法可以较好的均衡负载。

加权最少连接

主要用于集群中服务器性能差异大的情况下，调度器可以优化负载性能，具有较高权值的服务器可以将承受较大的活动连接。

基于局部性的最少连接

主要是针对目标IP地址的负载均衡，将请求的目标IP地址找到离其最近的服务器进行使用，如果服务器不存在或者满载的话，就会继续寻找下一个服务器。

带复制的基于局部性的最少连接

主要是针对目标IP地址的负载均衡，根据请求的目标IP地址找出该地址所对应的服务器，如果服务器不存在或者满载的话，就会继续寻找下一个服务器。当服务器有一段时间没有被修改，则会从最忙的服务器组中删除。

目标地址散列

根据请求的目标IP地址从静态分配的散列表中超出对应的服务器，如果找到可用的服务器且没有满载，则返回空。

根据请求的源IP地址从静态分配的散列表中超出对应的服务器，如果找到可用的服务器且没有满载，则返回空。

DNS轮询

DNS轮询是最简单的负载均衡方式。以域名作为访问入口，通过配置多条DNS记录使得请求可以分配到不同的服务器。DNS轮询没有快速的健康检查机制，而且只支持WRR的调度策略导致负载很难“均衡”，通常用于要求不高的场景。并且DNS轮询方式直接将服务器的真实地址暴露给用户，不利于服务器安全。

CDN

CDN（ContentDeliveryNetwork，内容分发网络）。通过发布机制将内容同步到大量的缓存节点，并在DNS服务器上进行扩展，找到离用户最近的缓存节点作为服务提供节点。因为很难自建大量的缓存节点，所以通常使用CDN运营商的服务。目前国内的服务商很少，而且按流量计费，价格也比较昂贵。

本文由百家号作者上传并发布，百家号仅提供信息发布平台。文章仅代表作者个人观点，不代表百度立场。未经作者许可，不得转载。