

基于一致性哈希的高可用负载均衡方案的设计与实现

林德煜

(中移互联网有限公司,广东 广州 510653)

摘要:负载均衡服务已经逐渐成为了限制整体应用集群可用性的重要内容,基于此,本文依托于一致性哈希,提出高可用的负载均衡方案。

关键词:服务集群;负载均衡;高可用性;一致性哈希

[DOI]10.12231/j.issn.1000-8772.2021.06.317

1 引言

为了在最大程度上保障服务的稳定性和可用性,要求相关服务器做出积极快速的响应,并针对用户请求予以及时处理。依托于服务器的横向拓展能力,让各项业务系统得以实现快速伸缩,进而提升节点的可用性,让服务器得以实现全天候运营。为切实满足上述需求,可以通过强化单个服务器配置和将多台服务器构成集群等方式予以解决。

2 负载均衡和高可用关键技术概述

2.1 负载均衡

为了提供具有更高性能和可用度的应用服务,一般通过服务集群的方式开展。由于不同服务器之间的硬件配置存在不同程度的差异,为了实现对于差异资源的充分利用,要求结合服务器的实际配置情况,承担用户所提出的相应需求,以免出现服务器资源分配不均的问题,让服务器集群的整体性能得到充分保障。为达成这一目标,要求依托于稳定可行的负载均衡算法,让后台服务器可以获取来自用户的各类请求信息,并将其予以合理分配,以便各个服务器均衡地承担此类任务,让服务器集群的吞吐率和可用性得到切实提升。

2.2 高可用

高可用性,也即一旦应用环境中的某一组件发生了故障问题,无论这一故障是基于何种层面,都要求其具备可以持续提供服务的能力。最为理想的情况就是可以确保机器在停机时,服务器的停止、切换和重启都可以保持在一个完全透明的状态之下。而一旦某一组件由于故障问题而难以持续服务,则需针对这一组件予以故障切换^[1]。

3 基于一致性哈希的高可用负载均衡方案设计

3.1 注册服务器

对于正向代理而言,一般借助客户端实现从服务请求到真实服务器地址的转换,并将其转化为 DNS 映射的负载均衡方案。在进行这一方案设计时,为了维持良好的互联网秩序,实现对于 DNS 解析技术的高效利用,需要依托于特定的方式,同时,为其提供充足的传播时间,方可实现对于服务集群的充分变更,并在其中建立相应的 IP。同时,在客户端之中针对 IP 解析请求予以缓存,避免出现经常访问的情形,并将其应用于 IP 解析服务器之中,用作监控服务器集群,以获取相应的机器地址列表。

3.2 一致性哈希映射

一致性哈希算法是由 Karger 在 1997 年所提出的一种算法结构,可以有效解决因特网之中的各类热点问题。通过一致性哈希算法,可以充分修正简单哈希算法之中所存在的各类问题,以便在 P2P 环境之中实现对于 DHT 的高效利用。简单而言,一致性哈希算法便是将由哈希值所构成的整体空间建构成为一个具有虚拟性质的圆环,以哈希函数 H 为例,如其哈希值为一个具有 32 位的无符号整形,则可以结合下图 1 展开对于其哈希空间环的表征。一致性哈希算法往往表现为数据分布的形式,但是仍然需要将数据存储于机器之中,为此,需要将机器映射到相应的环中,也即一致性哈希映射。在获取与集群相关的 IP 地址后,客户端所面临的一个重要问题就是怎样依托于某一策略,确定具体的服务请求地址,以实现对于客户端请求的均匀分布,让系统的负载均衡性能和拓展性得到充分保障^[2]。

4 基于一致性哈希的高可用方案的实现与部署

4.1 改进后负载均衡高可用方案的实现

第一,采集汇报模块,采集来自本地服务器的信息,并将其发送到注册服务器之中,将其安装于服务端上,以均衡其所在节点。第二,信息

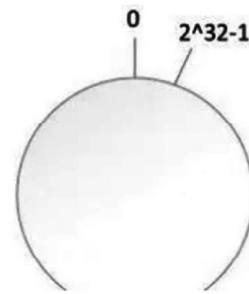


图 1 哈希环图示

注册模块,结合其所发送的信息,将其存储于存活列表之中,结合 CPU 和内存信息确定与节点相对应的权重,将其安装于注册服务器之中。第三,客户端插件,从注册服务器之中获取相应的负载信息和集群存活信息,达到一致性哈希环,结合相应请求返回服务节点的 IP。第四,客户端应用程序,可以以移动 APP 或后台服务的方式存在,是高可用方案的一部分,也是用户原生的应用代码。

4.2 负载均衡高可用方案与 OpenStack 的集成

上文已经提出了一种依托于一致性哈希环的服务高可用方案,可以将其搭载与负载均衡服务器之中,以提升负载均衡服务的高可用性,并将其应用于 OpenStack 云平台中,以构建相对稳定的方案。通常而言,需要相应管理员为各自的服务配置负载均衡器,为云平台租户进行负载均衡器的合理配置,并将其作为云平台的一个基础性组件予以处理。

5 性能测试与分析

5.1 测试环境

依托于 OpenStack 云平台,实现负载均衡高可用,并将测试环境分成 OpenStack 云平台和负载均衡组件两个部分。

5.2 测试案例与结果分析

5.2.1 高可用性测试。第一,依托于上文所述架构予以运行,并在合适的时间停掉传统方案的主节点和优化方案的任一节点。待运行一段时间后,恢复停止的机器,并针对其相应时间予以观察。在虚拟 IP 切换时,传统的负载方案会存在一定时间的无响应状态;至于改造后的高可用方案,只要集群之中始终保持存活的节点,便可以维持正常的工作状态。

5.2.2 负载测试。针对优化后的负载均衡高可用方案予以测试,明确其负载分布情况,结合相应节点,通过具有正常相应的服务 IP 实现对于负载均衡器处理效率的统计,并予以对比^[3]。

6 结束语

为了有效应对日益增长的业务需求,降低运行成本,要求针对不同服务器个体予以集群化建设,并在各个集群之中配置相应的负载均衡服务。随着负载均衡部署模式的持续变化,其可用性问题的重要性也得到了彰显,笔者针对负载均衡高可用问题做了相应探索,希望可以推动集群负载均衡服务的发展。

参考文献

- [1]于波,李庆建,王卫.实时通信集群负载均衡策略研究与应用[J].计算机系统应用,2019,28(05):167-172.
- [2]江淞,王宝海,赵金城,等.基于一致性哈希算法和 Ckafka 技术的 IMS 电话实时录音系统[J].计算机系统应用,2020,029(005):88-93.