

milkyty

芝兰生于深谷，不以无人而不芳。君子修身养德，不以穷困而改志。

博客园

首页

新随笔

联系

管理

使用nginx作为HTTP负载均衡

介绍

跨多个应用程序实例进行负载均衡是一种常用的技术，它被用于优化资源的利用率，最大化吞吐量，减少延迟并确保容错配置。

可以使用nginx作为一个非常有效的HTTP负载均衡器，将流量分配给多台应用服务器，并通过nginx提高Web应用程序的性能，可扩展性和高可靠性。

负载均衡方法

在nginx中支持以下负载平衡机制（或方法）：

- √ round-robin（轮询） — Nginx以轮询的方式对收到的请求进行分发。
- √ least-connected（最少连接） — Nginx会把下一个连接分配给具有最小活动连接的服务器。
- √ ip-hash（IP绑定） — 使用hash函数去决定下一个连接应该分配给哪一台服务器（基于客户端的IP地址）。

默认负载均衡配置

使用nginx做一个最简单的负载均衡配置可能如下所示：

- ```
1. http { //协议
2. upstream myapp1 { //服务器群
3. server srv1.example.com;
4. server srv2.example.com;
5. server srv3.example.com;
6. }
7. server {
8. listen 80; //监听端口
9. location / {
10. proxy_pass http://myapp1; //对服务器群进行负载均衡
11. }
12. }
13. }
```

在上面的例子中，服务器srv1-srv3上运行了三个具有相同应用程序的实例. 当负载均衡方法没有进行特别的配置的话，默认为轮询。这时所有的请求都被nginx代理到服务器群组myapp1中，同时nginx使用HTTP负载均衡来对请求进行分发。

nginx的反向代理可以实现包括HTTP，HTTPS，FastCGI，uwsgi，SCGI和memcached在内的负载均衡

如果想配置HTTPS协议的负载均衡而不是HTTP协议的负载均衡，只需要使用"https"作为协议即可

当为FastCGI，uwsgi，SCGI或memcached配置负载均衡时，分别使用fastcgi\_pass，uwsgi\_pass，scgi\_pass和 memcached\_pass 指令。

### 最小连接负载均衡

#### 公告

昵称：milkyty  
园龄：6年6个月  
粉丝：91  
关注：81  
[+加关注](#)

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

#### 搜索

#### 我的标签

性能测试(95)

linux(52)

java(50)

数据库(46)

LR(41)

linux命令(31)

mysql(23)

jmeter(21)

jvm(20)

另一种负载均衡机制是基于最小连接的。在一些请求需要更长的时间才能完成的情况下，最小连接机制可以更加公正的控制应用程序实例的负载均衡。

使用最小连接的负载均衡，nginx将尽量不会把过多的请求来使忙碌的应用程序服务器超负荷运行，而是将新的请求分发到不太忙的服务器。

当使用 least\_conn指令作为服务器组配置的一部分时，将激活nginx的最小连接负载均衡：

```
1. upstream myapp1 { //配置服务器群组
2. least_conn; //最小连接机制命令
3. server srv1.example.com;
4. server srv2.example.com;
5. server srv3.example.com;
6. }
```

会话持久性

请注意，通过轮询或最小连接的负载均衡，每个后续客户端的请求都可能会被分配到不同的服务器。在这里并不能保证同一个客户端始终指向同一个服务器。

如果需要将客户端绑定到特定的应用服务器—换句话说，使客户端与nginx的会话进行“粘滞”或者“持久化”，就总是尝试去选择一个特定的服务器而言，ip-hash负载均衡机制在这里可以被使用。

使用ip-hash 机制，客户端的ip地址被用作hash表里的键，nginx根据键来确定应该为客户端请求选择服务器群组里的哪一个服务器。此方法会确保从同一个客户端发来的请求会始终定向到同一台服务器，除非当这台服务器出现不可用的时候。

要配置ip-hash负载均衡，只需将ip-hash指令配置到服务器（upstream）组的配置中。

```
1. upstream myapp1 {
2. ip_hash;
3. server srv1.example.com;
4. server srv2.example.com;
5. server srv3.example.com;
6. }
```

加权负载均衡

还可以通过使用服务器权重进一步影响nginx负载均衡算法。在上面的示例中，服务器权重没有配置，这就意味着所有指定的服务器被看做对特定负载均衡方法具有同等资格。

特别是轮询使用的时候，这也意味着只要有足够多的请求，并且以统一的方式处理请求并且完成速度足够快，这时 通过服务器进行的请求分配或多或少概率是相等的。

当为服务器的配置指定权重参数时，权重将会作为负载均衡决策的一部分进行计算。

```
1. upstream myapp1 {
2. server srv1.example.com weight=3; // 为此服务器增加权重
3. server srv2.example.com;
4. server srv3.example.com;
5. }
```

通过这种配置，每五个新的请求将会通过应用实例进行分发，如下所示：3个请求将会被分发给srv1,1个请求将会分发给srv2，另一个请求分发给srv3。

在最近的nginx版本中，同样可以在最小连接和ip-hash机制中使用负载均衡的权重。

健康检查

nginx 中的反向代理实现包括内带（或被动）的服务器运行状况检查。如果特定服务器的响应失败并出现错误，nginx则会把这台服务器标记为失效，并会尝试在一段时间内避免后续入站的请求选择此服务器。

该max\_fails 指令设置在fail\_timeout期间与nginx服务器连续的不成功的尝试连接的通信数量。默认情况下，max\_fails设置为1，当设置为0时，该服务器的运行状况检查会被禁用。fail\_timeout参数还定义了服务器将被标记为失效的时间。在 服务器发生故障后的fail\_timeout间隔之后，

nginx将开始以实时客户端的请求优雅地探测服务器。如果探测成功，则将服务器标记为活动的。

文件上传下载

中间件相关(17)

更多

随笔分类

java(1)

Jmeter(6)

LR(3)

办公(1)

随笔档案

2018年4月 (1)

2018年2月 (2)

2017年11月 (2)

2017年9月 (1)

2017年8月 (2)

2017年7月 (5)

2017年6月 (3)

2017年5月 (2)

2017年4月 (2)

2017年2月 (1)

2016年11月 (1)

2016年10月 (1)

2016年9月 (6)

2016年8月 (3)

2016年6月 (2)

2016年5月 (3)

2016年3月 (2)

2015年11月 (1)

如果实现了负载均衡，除了Session问题，还会碰到文件的上传下载问题。文件不可能上传不同的服务器上，这样会导致下载不到对应文件的问题。我们看一下下面的方案

独立文件服务器  
文件压缩数据库

两种方案都是常用的，我们来说一下文件压缩数据库，以前的方式都是将文件二进制压缩至关系型数据库，而现在NOSQL的流行，加上MongoDB处理文件又比较方便，所以文件压库又多了一种选择。毕竟文件服务器的效率和管理以及安全都不及数据库。

nginx+tomcat集群中，使用Nginx session sticky实现。基于cookie的session机制。

Linux集群主要分成三大类( 高可用集群， 负载均衡集群，科学计算集群)。其实，我们最常见的也是生产中最常接触到的就是负载均衡集群。

负载均衡实现

- DNS负载均衡，一般域名注册商的dns服务器不支持，但博主用的阿里云解析已经支持
- 四层负载均衡(F5、LVS)，工作在TCP协议下
- 七层负载均衡(Nginx、haproxy)，工作在Http协议下

分布式session

大家都知道，服务一般分为有状态和无状态，而分布式sessoion就是针对有状态的服务。

分布式Session的几种实现方式

- 基于数据库的Session共享
- 基于resin/tomcat web容器本身的session复制机制
- 基于oscache/Redis/memcached 进行 session 共享。
- 基于cookie 进行session共享

分布式Session的几种管理方式

Session Replication 方式管理 (即session复制)

- 简介：将一台机器上的Session数据广播复制到集群中其余机器上
- 使用场景：机器较少，网络流量较小
- 优点：实现简单、配置较少、当网络中有机器Down掉时不影响用户访问
- 缺点：广播式复制到其余机器有一定延时，带来一定网络开销

Session Sticky 方式管理

- 简介：即粘性Session、当用户访问集群中某台机器后，强制指定后续所有请求均落到此机器上
- 使用场景：机器数适中、对稳定性要求不是非常苛刻
- 优点：实现简单、配置方便、没有额外网络开销
- 缺点：网络中有机器Down掉时、用户Session会丢失、容易造成单点故障

吃透这套架构演化图，从零搭建Web网站也不难！

2015年8月 (2)

2015年6月 (1)

2015年3月 (1)

2014年10月 (1)

2014年9月 (1)

2014年8月 (1)

2014年5月 (1)

2014年3月 (1)

2014年2月 (5)

2014年1月 (13)

2013年12月 (8)

文章分类

android(1)

java web(5)

linux(9)

python(1)

笔记(1)

测试工具(8)

持续集成

数据结构与算法(1)

友情链接

beigai

hualusiyu

itpub

lifetragedy

MKing's Blog

标签: 系统架构

好文要顶

关注我

收藏该文

milky

关注 - 81

粉丝 - 91

+加关注

« 上一篇: Nginx服务状态监控

» 下一篇: Nginx配置说明

posted @ 2017-09-12 09:32 milky 阅读(55) 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:

· nginx作为tcp负载均衡

· 【翻译】Nginx的HTTP负载均衡

· Nginx作为负载均衡器upstream

· 使用Nginx实现负载均衡

· 使用nginx实施负载均衡

最新新闻:

· 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”

· 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

» 更多新闻...

|                                  |
|----------------------------------|
| PrefTest性能测试工作室                  |
|                                  |
| <a href="#">不胜人生一场醉</a>          |
|                                  |
| <a href="#">郭亨的博客</a>            |
|                                  |
| <a href="#">酷壳CoolShell - 陈皓</a> |
|                                  |
| <a href="#">蓝冰咖啡</a>             |
|                                  |
| <a href="#">宋运剑</a>              |
|                                  |
| <a href="#">淘测试</a>              |
|                                  |
| <a href="#">行知</a>               |

|                                                                      |
|----------------------------------------------------------------------|
| 最新评论                                                                 |
|                                                                      |
| 1. Re:linux命令——II                                                    |
|                                                                      |
| @叶琦彰这个就是文件的<br>吧...                                                  |
|                                                                      |
| 2. Re:linux命令——II                                                    |
|                                                                      |
| 应该是user而不是owne                                                       |
|                                                                      |
| 3. Re:java命令--jstack 二                                               |
|                                                                      |
| 推荐:                                                                  |
|                                                                      |
| 4. Re:jmeter学习记录--生成报告                                               |
|                                                                      |
| 在win7环境下，执行-e ·<br>果index.html页面，jme<br>x -l result.jtl -e -o /tmp1: |
|                                                                      |
| 5. Re:redis cli命令                                                    |
|                                                                      |
| 十分详细                                                                 |