

nginx中实现负载均衡调度的模块ngx_http_upstream_module模块，nginx是一个非常高效的HTTP负载均衡器。负载均衡的指令只能配置到httpd段。

示例配置

```
1 upstream backend {
2     ip_hash;
3     server 192.168.1.151 max_fails=number
fail_timeout=10s;
4     server backend2.example.com:8080;
5     server unix:/tmp/backend3;
6
7     server backup1.example.com:8080    backup;
8     server backup2.example.com:8080    backup;
9 }
10 server {
11     location / {
12         proxy_pass http://backend;
13     }
14 }
```

调度方式

1. 轮询(round-robin)。轮流将请求调度到后端服务器，nginx负载均衡器的默认调度算法。
2. 最少连接(least_conn)。其中请求被传递到具有最少数量的活动连接的服务器，同时考虑服务器的权重。如果有几个这样的服务器，则依次使用加权循环平衡方法。
3. 会话持久(ip-hash)。使用ip-hash时，客户端IP地址作为hash key使用，用来决策选择服务器集群中的哪个服务器来处理这个客户端的请求。这个方法保证从同一个客户端发起的请求总是定向到同一台服务器，除非服务器不可用。

配置指令

server指令参数

1、weight

- 1 调度器将按照指定的权重比例调度到后端服务器。默认的权重的1
- 2 `weight=5;`

2、max_fails

设置健康监测次数，即监测几次失败后，停掉被代理服务器，默认为1

`max_fails=3;`

3、fail_timeout

默认为10分钟。负载均衡器尝试与后端服务器通信max_fails指定的次数都失败后，之后的fail_timeout时间段

内都将不在进行通信，超过fail_timeout时间段后，在进行max_fails次数尝试连接，如果失败，则断定为不可达服务器，之后不在尝试连接。

`fail_timeout=60s;`

4、backup

标记为备用服务器

5、down

标记停掉某台服务器

ip_hash

指定使用负载均衡算法，基于客户端ip地址作调度。

least_conn

指定使用负载均衡算法，最少连接。