服务器压力测试

## 编写目的

采用轻量级压力测试方式，检测服务器对高并发浏览的响应能力

## 工具

WRK是一个现代的HTTP基准测试工具，能够在单个多核CPU上运行时产生显着的负载。它将多线程设计与可扩展事件通知系统结合。

## 测试关注点

在开始测试之前，我们需要了解一下系统吞吐量（TPS）、用户并发量、性能测试概念和公式，以及这三个参数对我们服务器响应能力的影响。我们关心的也大多都是并发量，TPS,和响应时间这3者，WRK中也正好有对应三者的测试结果。

一个系统的吞度量（承压能力）与request对CPU的消耗、外部接口、IO等等紧密关联。

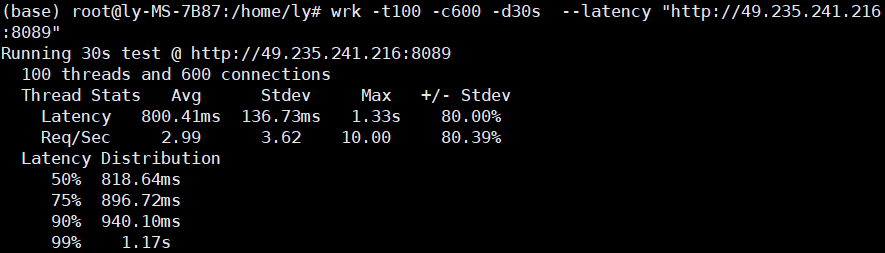
## 开始测试

首先我们需要定下目标，和环境，目前我使用安装在AMD Ryzen5 2400G上的Ubuntu 18.04 LTS平台含有16G内存做测试机，服务器部署在Tencent cloud单核心2G内存的机器上出口带宽仅有1Mbps。

基于现有服务器资源的局限性，我们需要在用户打开app时与服务器通信并下载登陆所需验证码，验证码为JPG格式大小约为5KB，在完成登陆以后需要与用户同步Todo列表数据，列表数据平均大小约为1KB。按照测试计划标准尺度，要求服务器响应时间小于2s，并要求服务器通过30s千并发测试。30秒内完成千并发每条6KB以上数据传输，需要数据带宽约为0.8Mbps，这就要求服务器在保证响应时间同时跑满出口带宽才能完成测试

## 一轮测试

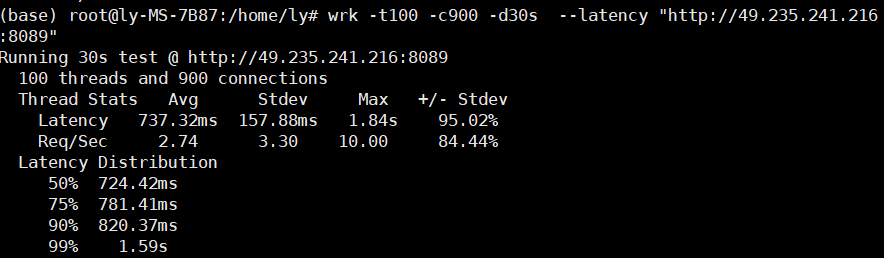
先进行600并发测试



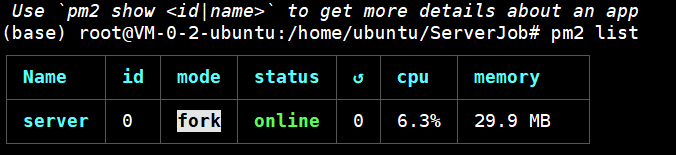
可以看到在600并发量下，持续测试30s的时间一共产生了将近10万个请求，平均响应时间是800ms，且最大响应时间没有超过1.5s，但存在3个timeout请求

## 负载均衡测试

采用PM2系统进行负载均衡操作，并在近千并发量下对服务器进行压力测试



从测试结果可以看出，采用PM2进行负载均衡后在并发数提高情况下平均响应时间有所降低，但最大响应时间有较大提升。通过PM2查看测试时服务器资源使用情况，可以看到CPU使用率与内存占用均不是很高。



## 测试总结

由于云服务器仅有一个中央处理器资源可以同时支持两个线程的数据处理。在使用负载均衡后，PM2对不同线程调度使得平均响应时间下降，但由于调度策略计算与执行本身需要占用系统资源，于是造成最大响应时间的提升。但总体来说，平均响应时间被控制在1s内，且timeout次数被控制在10以内。由此可得出结论次服务器可支持千并发请求，并保持1s以内的响应时间。