

## 性能测试常见术语浅析

之前在性能测试过程中，对于某些其中的术语一知半解，导致踩了很多坑。这篇博客，就常见的一些性能测试术语进行一次浅析。。。

### 负载

对被测系统不断施加压力，直到性能指标超过预期或某项资源使用达到饱和，以验证系统的处理极限，为系统性能调优提供依据；

### 并发

①狭义上的并发：所有用户在同一时间点进行同样的操作，一般指同一类型的业务场景，比如1000个用户同时登陆系统；

②广义上的并发：多个用户与系统发生了交互，这些业务场景可以是相同的也可以是不同的，交叉请求和处理较多；

### 压力

系统在一定饱和状态下，例如CPU、内存等饱和情况下，系统能够处理的会话能力，以及系统是否会出现错误

特点：主要目的是检查系统处于压力情况下应用的性能表现，重点在于有无出错信息产生，系统对应用的响应时间等

### 事务

性能测试中，事务指的是从端到端，一个完整的操作过程，比如一次登录、一次筛选条件查询，一次支付等；

### 吞吐量

指在一次性能测试过程中网络上传输的数据量的总和，也可以这样说在单次业务中，客户端与服务器端进行的数据交互总量；

对交互式应用来说，吞吐量指标反映服务器承受的压力，容量规划的测试中，吞吐量是重点关注的指标，它能够说明系统级别的负载能力，另外，在性能调优过程中，吞吐量指标也有重要的价值；

#### 吞吐量和负载之间的关系：

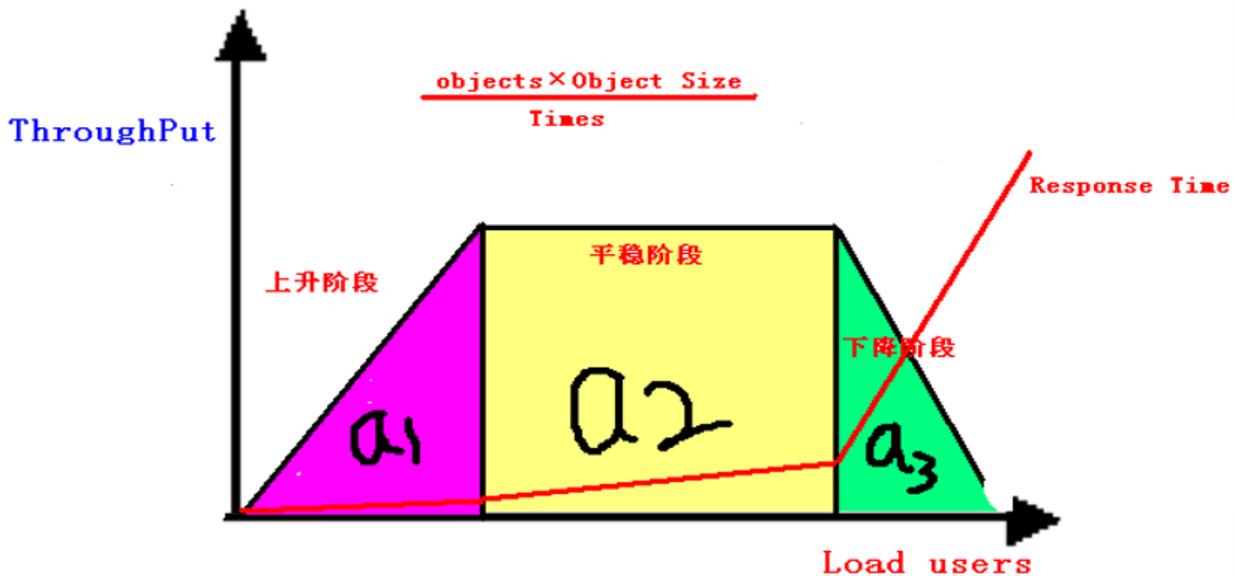
①上升阶段：吞吐量随着负载的增加而增加，吞吐量和负载成正比；

②平稳阶段：吞吐量随着负载的增加而保持稳定，无太大变化或波动；

③下降阶段：吞吐量随着负载的增加而下降，吞吐量和负载成反比；

**总结：**吞吐量干不过负载！！！！

如下图所示：



a1面积越大，说明系统的性能能力越强，a2面积越大，说明系统稳定性越好，a3面积越大，说明系统的容错能力越好（啧啧，图有点丑。。。）

## 吞吐率

吞吐量/传输时间，即单位时间内网络上传输的数据量，也可以指单位时间内处理客户请求数量，它是衡量网络性能的重要指标。

通常情况下，吞吐率用“字节数/秒”来衡量，当然，也可以用“请求数/秒”和“页面数/秒”来衡量；

## TPS

Transaction Per Second：每秒事务数，指服务器在单位时间内（秒）可以处理的事务数量，一般以request/second为单位；

PS：看到很多博客或性能测试人员将QPS和TPS混为一谈，个人认为，他们是以测试结果的统计得到该结论的；

QPS是查询，而TPS是事务，事务是查询的入口，也包含其他类型的业务场景，因此QPS应该是TPS的子集！

## QPS

Query Per Second：每秒查询率，指服务器在单位时间内（秒）处理的查询请求速率；

PS：TPS和QPS都是衡量系统处理能力的重要指标，一般和并发结合起来判断系统的处理能力；

## PV

Page View：页面浏览量，通常是衡量一个页面甚至网站流量的重要指标；

细分的话，有独立访问者数量、重复访问者数量、单独页面访问数量、用户停留时间等类型；

## RT/ART

Response Time/average Response Time：响应时间/平均响应时间，指一个事务花费多长时间完成；

一般来说，性能测试中平均响应时间更有代表意义。细分的话，还有最小最大响应时间，50%、90%用户响应时间等；

## Thinking Time

思考时间，在性能测试中，模拟用户的真实操作场景。用户操作的事务与事务之间是有一定间隔的，引入这个概念是为了并发测试（有交叉业务场景）时，业务场景比率更符合真实业务场景；

## 连接池

连接池是一个进程，多个连接在一个进程中存储、管理，它是共享、可复用的；

当客户端发起请求，先检查是否有闲置连接，如果有，则分配该连接给其使用；如果没有，则请求进入等待队列（等待空闲连接分配，这个取决于调度配置）或新建一个连接对象供其使用（取决于连接池有多少连接以及允许的最大连接数）；

每次客户端发起请求，如果都新建连接，会消耗很多的资源，连接池的存在及其特性，减少了连接的建立所消耗的资源以及节省了很多连接创建时间，给系统提供了更好的伸缩性，也有助于服务器性能的提升。

关于连接池中线程的状态，可参考此篇博客：

<https://my.oschina.net/cctester/blog/991744>

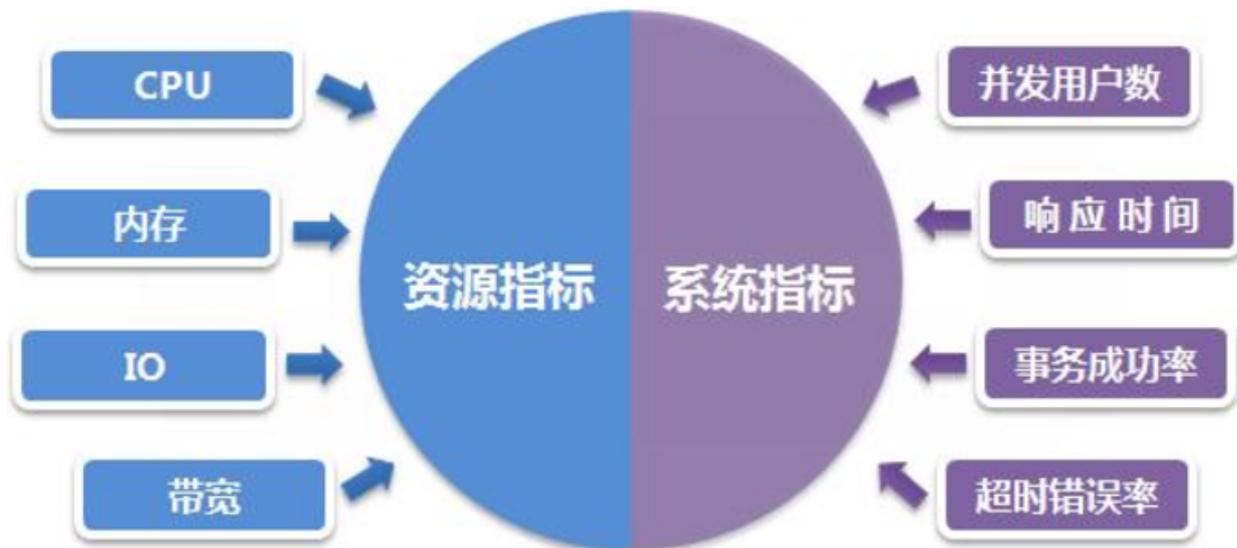
## 标准方差

各数据偏离平均数的距离（离均差）的平均数，它能反应一个数据集的离散程度。离散程度越大，数据越不可靠；

性能测试中引入这个术语，是为了对高峰期、平缓期的系统响应时间分布，不同业务场景的响应时间分布，以及I/O数在时间段上的分布等情况进行分析，以判断系统的稳定性。

## 资源使用率

如下图所示，资源指标与硬件资源消耗直接相关，而系统指标则与用户场景及需求直接相关：



### 资源指标：

CPU使用率：指用户进程与系统进程消耗的CPU时间百分比，长时间情况下，一般可接受上限不超过85%；

内存利用率：内存利用率=（1-空闲内存/总内存大小）\*100%，一般至少有10%可用内存，内存使用率可接受上限为85%；

磁盘I/O：磁盘主要用于存取数据，因此当说到I/O操作的时候，就会存在两种相对应的操作，存数据的时候对应的是写I/O操作，取数据的时候对应的是读I/O操作，一般使用% Disk Time（磁盘用于读写操作所占用的时间百分比）度量磁盘读写性能；

网络带宽：一般使用计数器Bytes Total/sec来度量，其表示为发送和接收字节的速率，包括帧字符在内；判断网络连接速度是否是瓶颈，可以用该计数器的值和目前网络的带宽比较；

**系统指标：**

并发用户数：单位时间内与系统发生交互的用户数；

在线用户数：某段时间内访问系统的用户数，这些用户并不一定同时向系统提交请求；

平均响应时间：系统处理事务的响应时间的平均值；事务的响应时间是从客户端提交访问请求到客户端接收到服务器响应所消耗的时间；

事务成功率：性能测试中，定义事务用于度量一个或者多个业务流程的性能指标，如用户登录、保存订单、提交订单操作均可定义为事务，单位时间内系统可以成功完成多少个定义的事务，

在一定程度上反应了系统的处理能力，一般以事务成功率来度量；

超时错误率：主要指事务由于超时或系统内部其它错误导致失败占总事务的比率；