

1) 负载：模拟业务操作对服务器造成压力的过程，比如模拟100个用户进行发帖。

2) 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；

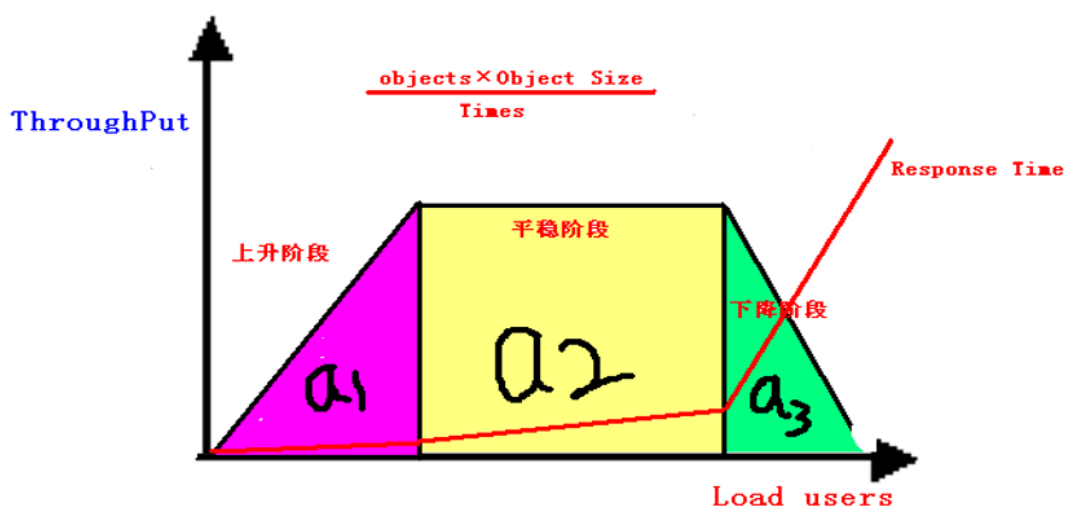
吞吐量和负载之间的关系：

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：吞吐量不会超过负载！！



a_1 面积越大，说明系统的性能能力越强， a_2 面积越大，说明系统稳定性越好， a_3 面积越大，说明系统的容错能力越好

3) 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；

4) 性能测试 (performance testing)：模拟用户负载来测试系统在负载情况下，系

统的响应时间、吞吐量等指标是否满足性能要求。（一般我们的压测指的就是这种，即能否满足基本性能要求。）

5) 负载测试 (load testing)：在一定软硬件环境下，通过不断加载负载来确定在满足性能指标的情况下能够承受的最大用户数。（根据cpu, mem, network, disk等使用情况，找出系统性能的拐点，给予生产环节规划建议）

6) 压力测试 (stress testing)：在一定软硬件环境下，通过高负载的手段来使服务器资源处于极限状态，测试系统在极限状态下长时间运行是否稳定。（相当于测试服务器环境稳定性。）

7) 稳定性测试 (Endurance testing)：在一定软硬件环境下，长时间运行一定负载，确定系统在满足性能指标的前提下是否运行稳定。（区别于压测在于，不比把系统资源调用到最高，在这满足基本性能指标即可。）

8) TPS：每秒完成的事务数，通常指每秒成功的事务数。（一个事务可以是一个操作，也可以是一些操作的集合。取决于你想测试的业务。）

9) 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。（通常ART更有意义，RT通常代表ART）

10) PV (Page View)：每秒用户访问页面的次数。（此参数用来分析平均每秒有多少用户访问页面）

11) Concurrency 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；

12) 思考时间 (Think Time)：模拟正是用户在实际操作时的停顿间隔时间。从业务的角度来讲，思考时间指的是用户在进行操作时。每个请求之间的间隔时间。（在脚本中应该体现为两个请求的间隔时间）

13) 标准差 (Std.Deviation)：根据数理统计的概念得来，标准差越小，说明波动越小，系统越稳定；反之，代表不稳定。（考量对象：RT、TPS、load、CPU等等）