

1.什么是软件性能

软件性能是个很大的概念，覆盖面非常广泛，对一个软件系统而言，包括软件执行效率、资源占用率、稳定性、安全性、兼容性、可扩展性、可靠性等等

- **用户视角**：针对用户来说，软件性能就是软件对用户的响应时间
- **管理员视角**：系统的响应时间、系统运行时服务器的状态，如cpu内存等使用情况、系统是否能够实现拓展、系统支持多少用户访问、系统性能的可能的瓶颈在哪里、系统是否能够7*24小时业务的访问
- **开发视角**：用户关心的响应时间、管理员关心的系统是否可扩展、架构是否合理、数据库是否合理、代码中是否存在性能问题、系统中是否有不合理的内存使用方式

2.什么是软件性能测试

性能测试是通过自动化的测试工具模拟多种正常、峰值以及异常负载条件来对系统的各项性能指标进行测试。**负载测试**和**压力测试**都属于性能测试，两者可以结合进行。通过**负载测试**，确定在各种工作负载下系统的性能，目标是测试当负载逐渐增加时，系统各项性能指标的变化情况。**压力测试**是通过确定一个系统的瓶颈或者不能接受的性能点，来获得系统能提供的最大服务级别的测试。（摘自百度百科）

3.什么系统需要做性能测试

web、app、单机应用程序（c/s）

4.性能测试的目的

- 评估系统的能力：测试中刚得到的压力水平和相应时间数据可以用于验证系统是否达到规划的水平
- 识别系统中的弱点：将系统的压力增加到一个极端水平，从而帮助发现系统薄弱环节并修复系统的瓶颈
- 验证系统的稳定性和可靠性：长时间的测试可能导致程序发生内存泄漏等引起的隐藏问题，在一个生产负荷下执行测试一定的时间，评估系统的可靠性是否达到满足
- 系统调优：重复执行性能测试，优化代码、配置等，以验证系统调优是否取得明显提升
- 系统扩容依据：设计系统增量的变化，为系统扩容方案提供数据依据

5.性能测试中有哪些常见的术语

并发数

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；

TPS/QPS

①TPS：Transactions Per Second，意思是每秒事务数，具体事务的定义，都是人为的，可以一个接口、多个接口、一个业务流程等等。**一个事务是指事务内第一个请求发送到接收到最后一个请求的响应的过程，以此来计算使用的时间和完成的事务个数。**

②QPS：Queries Per Second，意思是每秒查询率，是一台服务器每秒能够相应的查询次数，是对一个特定的查询服务器在规定时间内所处理流量多少的衡量标准。

TPs和Qps的区别

TPS：（Transactions Per Second），即每秒执行的事务总数。

首先一个事务包括三个动作，即客户端请求服务端，服务端内部进行处理，服务端对客户端进行响应。

将这三个动作看成一个整体，并将之称为一个事务，若在一秒内，服务端可以完成N个事务，则这个服务端的TPS为N。

一般来说，评价系统的性能主要看系统的TPS，系统的整体性能取决于性能最低模块的TPS值。

（木桶的容量取决于最短板，当然你不能把木桶斜着放，斜着放虽然能够改变木桶容量的依赖，此时依赖变为最长板，但会产生额外的消耗，比如需要一直扶住木板，或者产生木桶歪倒的风险）

QPS：（Queries Per Second），及每秒执行的查询总数（每秒有多少的请求响应）

客户端请求一个地址时，比如百度首页，其实会产生很多的请求，比如js、css、png等，像这样的每个单个请求都可以算作查询次数。

若在一秒内，客户端请求服务端的首页，服务端返回了N个内部链接（js、css、png、html等），那么服务端的QPS就为N。

QPS反映系统的吞吐能力，更偏向于读取文件，查询数据。

若在一秒内，用户请求了百度首页并看到了首页全貌，这样就形成了一个TPS，但却形成了多个QPS。

若在一秒内，我们请求一个单调的网页，此网页只有一个html，不包含任何其他内部链接，此时TPS=QPS。

思考时间

思考时间：：操作时的时间间隔，它可以模拟真实的用户操作，减少服务器压力，满足实际的业务规则，一般情况下加思考时间，但针对于目的只关注于系统性能而不考虑用户场景的可以去掉思考时间

资源利用率

资源利用率：资源利用率分为应用服务器的资源利用率和数据库服务器的资源利用率，一般监测的指标：CPU、LoadAverage、Memory、队列、IO、网络等

PV/UV

①PV：Page View，即页面浏览量或点击量，用户每1次对网站中的每个网页访问均被记录1个PV。用户对同一页面的多次访问，访问量累计，用以衡量网站用户访问的网页数量。

②UV：Unique visitor，访问您网站的一台电脑客户端为一个访客。00:00-24:00内相同的客户端只被计算一次

RT

RT：Response Time响应时间，指应用执行一个操作所需的时间，包括从发出请求开始到最后收到响应所需要的时间

连接池

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。

负载

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；

吞吐量

吞吐量是指系统在单位时间内处理请求的数量，TPS、QPS都是吞吐量的常用量化指标。

系统吞吐量要素

一个系统的吞吐量（承压能力）与request（请求）对cpu的消耗，外部接口，IO等等紧密关联。

单个request对cpu消耗越高，外部系统接口，IO影响速度越慢，系统吞吐能力越低，反之越高。

重要参数

QPS(TPS)、并发数、响应时间

QPS(TPS)：每秒钟request/事务数量
并发数：系统同时处理的request/事务数
响应时间：一般取平均响应时间

关系
$$QPS(TPS) = \text{并发数} / \text{平均响应时间}$$

一个系统吞吐量通常有QPS(TPS),并发数两个因素决定，每套系统这两个值都有一个相对极限值，在应用场景访问压力下，只要某一项达到系统最高值，系统吞吐量就上不去了，如果压力继续增大，系统的吞吐量反而会下降，原因是系统超负荷工作，上下文切换，内存等其他消耗导致系统性能下降。

6.性能测试的分类

1.基准测试

对新系统进行性能测试，记录各项性能指标，作为后续性能测试的基准。

2.负载测试

给系统定容定量，在系统上不断增加负载（用户数），看系统最大承受的用户负载是多少，看系统是否达到了需求所需要的性能指标，RT/TRP，吞吐率，资源利用率

3.压力（强度）测试

为了找出系统的性能瓶颈，以及产生系统性能瓶颈的原因。

一般是让系统的用户负载达到饱和和长时间运行，检查系统资源分配是否合理，同步问题，内存泄漏，检查原因。

4.并发测试

在一定的软硬件环境下，系统的其他条件（数据库数据容量不变），不断增加系统的用户数量。

（1）查看系统在不同级别的用户数量下，系统的指标的不同表现情况。

（2）查看系统的用户数量到达什么情况下，系统性能达到拐点

5.容量测试

在一定的软件环境下，系统的其他条件不变（用户数量不变），不断增加数据库的数据容量。

（1）查看系统不同级别的数据库容量下，系统性能指标的表现。

（2）找系统性能的饱和点（拐点）

6.配置测试

通过测试不同的应用服务器，网络环境，数据库不同的参数设置，不同硬件设备，查看系统的性能表现，找出最优最合理的资源配置。

7.可靠性测试

长时间运行系统，查看系统的性能指标（响应时间，TPS，点击率，吞吐量，资源利用率）是否稳定

（70%~90%）