

软件性能测试研究与分析

□陈 江/文

随着互联网的高速发展,软件的性能测试越来越受到软件生产商和用户的重视。2008 北京奥运会的举办成功,庞大的售票系统,若非没有经过一定程度的性能测试,是无法满足全世界人民共同购票需求的。一个网站从刚开始发展时只有几个人,随着用户的逐步增加,使用者可能会几十成百倍数量级的增长,通常情况下网站可能会崩溃,因此软件性能测试是非常重要和必要的。

性能测试的意义

所谓的性能是指表明软件系统或构件对于其及时性要求符合程度的指标。它属于软件产品的特性范围,一般用响应时间、吞吐量每秒点击数等参数指标来进行衡量。性能测试是一种规范,可以量化更改业务指标所产生的影响,进而说明部署的风险,因此性能测试是软件质量保证环节必不可少的环节。

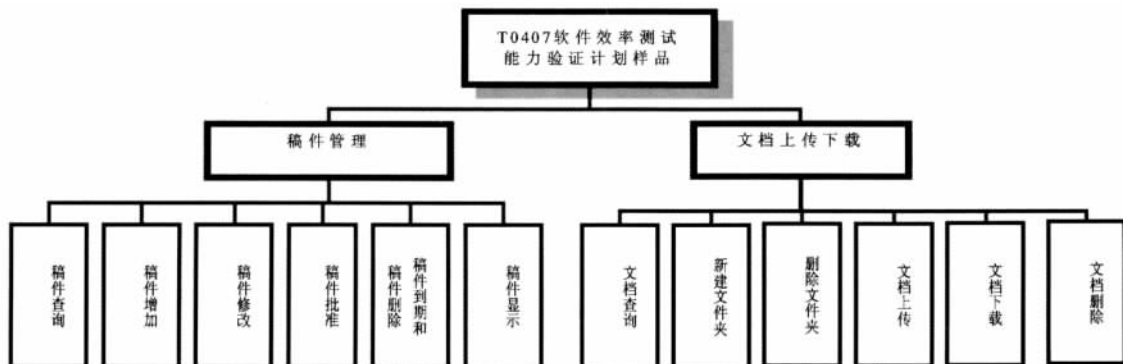
性能测试的划分

性能测试的定义可以从广义和狭义两个方面考虑。狭义的性能测试是指通过模拟生产的压力或场景来看是否满足生产性能的要求。广义上是指压力测试、负载测试、强度测试、并发(用户)测试、容量测试、配置测试、可靠性测试的统称。

案例实战

本文主要对案例进行分析和阐述测试工具(loadrunner9.0)在实际测试中的应用,并使用loadrunner对测试的分析结果进行总结,结合自身工作体会和经验和大家一起交流和学习。

【案例】2008年10月初国家应用软件产品质量监督检验中心负责实施中国合格评定国家认可中心(CNAS)的能力验证计划项目,并提供一个软件样品,要求各个CNAS联盟单位的软件评测中心独立进行该样品软件的效率测试,这里所提出的效率测试就是性能测试。该样品软件是一个网站稿件管理发布系统,只有两个相对独立的功能,即稿件管理和文档下载功能。主要功能如下图



具体性能要求：

- (1) 主要单点操作在 5 秒钟内完成；
- (2) 系统支持 50 人在线；
- (3) 稿件管理的主要功能至少支持 20 人并发访问；
- (4) 上传 512K 大小文档,速度达到 20 篇 / 分钟；
- (5) 系统可以连续稳定运行 12 小时。

该系统从架构上看是 B/S 模式，具体的硬件和测试端的硬件环境和软件环境这里就不做叙述，WEB 服务器 ip 地址为 192.168.8.8，操作系统是 windows server 2003。

(一)首先来考虑第一个性能指标：“主要单点操作在 5 秒钟以内完成”。一般来说，软件本身的单点操作非常多，像按钮单击事件基本都是单点操作，我在这里仅举一个例子来说明，具体测试参看下图，用 loadrunner 录制该步骤来获得脚本，得到测试后的结果：

用例编号	R08T0407_TC_01			
性能描述	单点操作响应时间			
用例目的	测试用户进行稿件查询所需要的响应时间			
前提条件	用户已登陆			
测试场景设计	一个虚拟用户、迭代时间间隔5秒、迭代次数100			
测试结果	基本查询	小于5秒	第一次	3.16秒
			第二次	0.41秒
			第三次	0.40秒

从上面的结果中，我们可以分析，最后实际测试的结果都在 5 秒以内，符合提出的性能指标。

(二)“系统支持 50 个人在线”。从这里可以看出我们应该做的是压力测试，用户要求能够让 50 个用户能够登录系统，也就是说数据库允许在线 50 个用户。那么我们具体的方法是用每隔一秒或几秒钟让用户不断上线，用 loadrunner 可以实现这样的过程，结果如下表：

用例编号	R08T0407_TC_02		
性能描述	系统支持50人在线		
用例目的	测试系统能够达到的在线人数		
测试场景设计	50个虚拟用户登录，运行30分钟。		
测试结果	被测对象	期望的在线人数	实际的在线人数
	在线人数	50	50以上

上面的表格中提到的数据池是指有 50 个不同用户名的用户的数据，loadrunner 模拟了 50 个不同的用户分时地登录该系统，在运行的时候不断从数据池里读取不同用户名的数据，运行脚本 30 分钟，最后结果显示该系统能够支持 50 个用户在线。当然可能还可以支持更多的用户在线，但是用户提出的技术指标已经达到，我们没有必要去测试该系统的极限。

(三)“稿件管理的主要功能至少支持 20 人并发访问”。这里就是典型的并发（用户）测试，具体的方法是用 loadrunner 录制创建脚本，在脚本里面插入集合点，集合点的意思就是指多个用户同时运行脚本到达这个集合点，然后虚拟用户同时开始做并发操作，看看该事务是否能够正常实现，下面列举了 20 个用户并发访问的操作，分析结果如下：

用例编号	R08T0407_TC_03			
功能描述	稿件查询操作至少支持20个用户并发访问			
测试场景设计	20个用户并发，观察事务执行情况以及web服务器的主要性能计数器。			
并发用户数	最小响应时间	平均响应时间	最大响应时间	事务成功率
20	0.41	2.29	3.22	100%
	CPU使用率	内存的可用数	磁盘I/O	
	<39	>1455，平均1464	平均4.9	

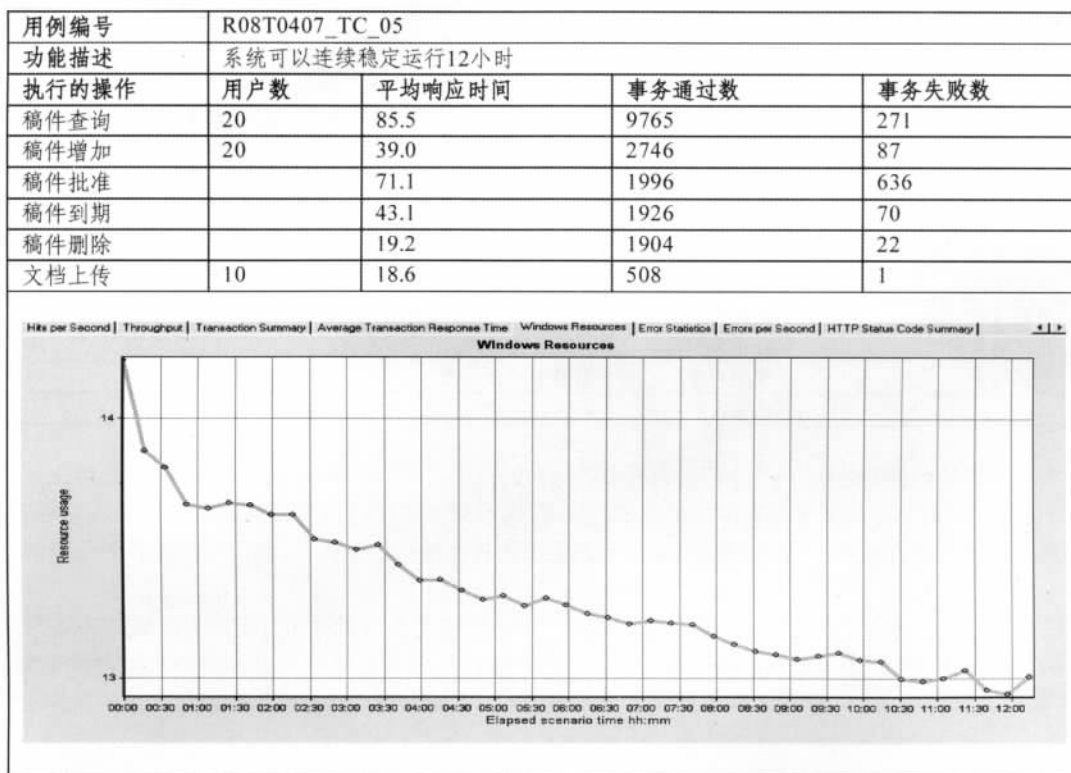
由于指标提出的是能够支持 20 用户并发,可以看出整个加压的过程所能够所体现出来的平均响应时间的变化,事务成功率都为 100%,说明系统资源远远未达到瓶颈,参考 cpu 的使用率、内存的可用数、磁盘 I/O 的数据中就可以看出来。

(四)“上传 512K 大小文档,速度达到 20 篇/分钟”。从指标上我们可以看出需要做容量测试,针对这个软件的指标本身,这个测试要求是做吞吐量测试,具体的方法和上面的一样,也使用测试工具 loadrunner 进行录制脚本和分析,具体的步骤和结果如下表:

用例编号	R08T0407_TC_04		
性能描述	上传 512K 大小文档,速度达到 20 篇/分钟		
用例目的	测试系统处理文档上传操作的吞吐量		
测试结果	并发用户数	期望的性能(平均值)	实际性能(平均值)
	20	大于 20 篇/分钟	57 篇/分钟

从结果我们可以看出,做容量测试是与并发相结合一种测试,从容量测试的概念可以完全体现出来,当然我们做此吞吐量的测试选择了不同用户数量的并发,只是在控制用户加压的方法不同,当然并发用户数越多,实际的吞吐量开始呈递减趋势,这也是正常的可以预料的结果。在用户步骤中,由于多个用户并发产生的上传文件将产生多个重复的文件名,因此这里做个技术处理,让不同并发用户每个产生的上传文件的文件都不同,避免出现重名的情况。

(五)“系统可以连续稳定运行 12 小时”。从字面上我们可以很轻松地看出来,这个指标是指系统的可靠性。系统的可靠性可以说是关系到系统软件的生命质量,可靠度越高,质量就越好,因此在做软件性能测试的时候一般都需要考虑到这个问题。有的系统可能在你做了无数次其他指标的测试,都不会出现问题,但是往往在可靠性方面的级别太低,所以绝对不能忽视可靠性这个指标,一些比较重要的数据可能平时正常使用不会出问题,但是一旦可靠性不高,就有可能是灾难性的大事故。做可靠性测试的时候一般原则上都要监视系统所在平台硬件的一些关键信息,如 windows resources、内存、cpu 等使用率的情况。



上面图表表明使用 loadrunner 进行多个事务的运行,然后连续运行 12 小时,分析结果是该系统存在一定程度的内存泄漏,这个也是不能忽视的问题。还有一个是事务失败率,由于用户没有提出具体要求的事务失败率应小于多少的指标,所以在这里我仅列出数据供用户方参考。

以上做了性能测试有关测试类别的分析和研究,但这只是其中冰山一角,总的说来,使用测试工具对软件系统进行分析 and 测试,有许多方法和尝试,其中技巧和因软件本身性质(如:多个协议支持)、复杂程度而有所不同,并没有什么固定的模式和方法。

(作者单位:福建省电子产品监督检验所福建省软件评测中心)