

Web 应用性能测试进展^{*}

芮素娟^{1,2} 丁晓明^{1,2}

(西南大学计算机与信息科学学院 重庆 400715)¹

(重庆市智能软件与软件工程重点实验室 重庆 400715)²

摘要 Web 性能测试是 Web 应用开发过程中的一个关键环节,然而 Web 站点的复杂性及其用户行为的不可预见性使得对其性能测试极其困难,大多数 Web 站点性能测试不精确并且不现实,不能够反映真实情况的性能测试用处不大或者是容易引起误导。本文主要介绍了 Web 站点性能测试的重要性、概念及其流程,国内外研究的现状,介绍了性能测试时用到的几个重要的指标,最后总结全文并进行了展望。

关键词 负载测试,性能测试,Web 应用

Progress in Web Applications Performance Testing

RUI Su-Juan^{1,2} DING Xiao-Ming^{1,2}

(Faculty of Computer and Informances Science, Southwest China University, Chongqing 400715)¹

(Intelligent Software and Software Engineering Research Institution, Chongqing 400715)²

Abstract Web performance testing is a crucial step in the Web application development. However, the complexity of Web and unpredictable operation of its users make it very difficult to test the performance of Web. As a result, most performance testings involved are inaccurate and impractical, which fail to reflect the real condition sometimes even be misleading. This thesis first explains the importance, concept and procedure of web performance testing; also introduces the current study in the field. Then it presents some important measures in the progress and at last summarize the thesis with a view of future.

Keywords Load testing, Performance testing, Web application

1 引言

Web 正以其广泛性、交互性、快捷性和易用性等特点越来越受到企业和个人的青睐,越来越多的商家使用 Web 站点来提供他们的产品,这样可以在降低成本的同时增加销售额。Web 应用软件的规模不断扩大,复杂性也逐渐增加,随之而来的问题就是 Web 站点流量的迅速增长,导致一些经常使用的 Web 站点由于负载过重而变得反应迟缓。当站点的性能问题达到无法忍受的程度时,就会导致用户过早地终止其在这个网站的事务。对于商业网站来说,用户放弃他们的网站将导致收入的减少。据估计,电子商务网站因性能不好而导致每年有接近 43.5 亿美元的损失^[1]; Forrester 的分析家 Harley Hanning 指出,现在电子商务网站的浏览者到购买者的转化率大概只有 2%,一个最大的原因就是人们不能容忍所访问站点的糟糕性能,例如页面不可访问或者下载时间太长。对于网上冲浪的一组人员来说,他们对很长时间的下载时间表现得十分不耐烦,他们希望在很短的时间内获得更多的信息。因此,Web 性能测试是 Web 应用开发过程中的一个重要环节,通过 Web 应用的性能测试,尽可能模拟现实用户环境以保证 Web 应用的性能。

2 Web 性能测试的概念

性能测试是一种信息的收集和分析过程,过程中收集的数据用来预测怎样的负载水平将耗尽系统资源。然而,现实性能指标很难收集,性能问题是很复杂的,大多数 Web 站点的性能问题都取决于在改变的环境下,Web 站点在用户高峰时期的负载情况。

性能测试的类型通常有负载测试、紧张测试和忍耐测试

3 种^[2]。Web 站点的负载测试通过分析应用程序的性能需求,编制负载测试计划,构建测试环境,在不同的负载级别上按照编制的测试计划运行一组模拟用户行为的脚本来确定它可以达到的工作量,确定在不同负载下系统的性能,目标是测试当负载逐渐增加时,系统组成部分的相应输出项,例如吞吐量、响应时间、CPU 负载、内存使用等来决定系统的性能。通常是借助性能测试工具的帮助在模拟的环境中实施的,测试结束时,还要分析负载测试结果,根据结果达到优化系统的性能的目的。紧张测试则不考虑用户的思考时间,尽可能快地执行测试脚本,主要为了测试 Web 站点在最差的环境下接受超负荷的能力。在某些应用中,很有必要增加紧张测试以找到所测试的站点是否崩溃。忍耐测试是一种间隔时间较长的负载测试或者紧张测试,不同之处是,负载或紧张测试间隔仅维持几十秒,忍耐测试要延迟几小时甚至于几天。忍耐测试通常会发现一些莫名其妙的错误。忍耐测试对于很重要的应用或者是需要维持很长一段时间的应用是很重要的。

3 基于真实负载的性能测试的过程

Web 应用的开发和更新时间很短,性能测试周期相应地也很短,如何在有限的时间内完成负载测试计划是一个难度很高的挑战,为了提高 Web 应用的负载测试效率,就需要遵守一定的流程,这样测试起来才会有章可循,减少不必要的花费。负载测试的一般过程如图 1 所示。

1) 充分熟悉了解待测试软件,进行负载测试首先要对待测试软件有个全面的了解,分析它的性能要求,估计它的一般能够支持的最大的并发用户数目,高峰时期最多可以支持并发的用户数,如果使用测试人员来完成,则花费的人力物力都太大,所以要用测试工具来模拟虚拟用户,使用虚拟用户来生

^{*} 本文研究得到重庆市自然科学基金(CSTC2004BB0146)、重庆信息产业发展基金(200401021)的资助。

成 Web 应用程序的负载。

2) 根据从负载测试中提取的信息分析应用程序的响应时间,吞吐量等反映真实情况的指标,可以看到系统在所测试的负载等级下的表现,从而可以估计系统所能承受的负载等级大概在哪个范围之内。通过分析,可以发现应用程序的瓶颈信息,对系统的优化提供了前提。

3) 查看测试结果是否满足预期指标,如果满足,则结束测试,否则要进行系统调试,重新测试,这是一个反复的过程。

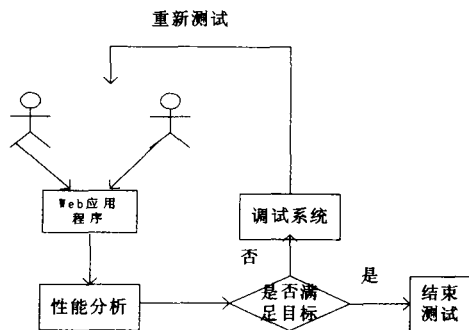


图1 基于真实负载的性能测试过程

4 性能测试研究现状

国外对性能测试的研究已经取得了许多成果,提出了一些性能测试方法,开发了相应的性能测试工具,主要如下:

1) 文[3]提出的一种负载测试方法,在文[14]中称之为虚拟用户方法,主要针对电子商务应用程序。通过网络进行交易可以增加销售和利润,但是它的成功要靠网络客户的支持,而只有让用户访问站点感到满意时,用户下次还会来访问此站点,相反,就会到竞争对手的网站去。

在软件发布之前,通过模拟真实用户的行为来对电子商务应用程序进行性能测试,可以预测在现实中可能会出现的情况,通过分析测试结果优化系统,以便系统能够及时地处理大量的负载,给用户提供良好的性能,在商业竞争中获取胜利。主要是通过性能测试工具或者手工模拟用户对应用程序施加压力,分析应用程序的响应时间等指标。这种方法有成熟的工具支持,使用性能测试工具可以用较少的硬件资源模拟很多个虚拟用户同时访问应用程序,还可以重现测试过程。

2) 基于网站使用签名(website usage signature, WUS)^[4],一个不能真实反映实际过程的负载测试是最没有用的^[5],并且它可能会产生误导作用,过高的估计 Web 站点的能力在实际的负载出现时会导致系统在发布之后崩溃,过低的估计将会产生不必要的延迟和成本。而 WUS 方法的提出是为了衡量测试负载和真实负载之间的接近程度,WUS 方法是一种有效的和系统的方法,能刻画网站的真实负载的一组关键变量和指标的集合。如果负载测试的 WUS 与实际的 WUS 不是太匹配,所测试的结论的有效性和可用性就要受到怀疑了,反之,如果它们很好的匹配,我们就可以放心的把测试的结论用到实际的情况中了。由于所用到的指标从网站实际运行所产生的日志文件中得到,适用于已经发布的程序,不适用于测试新开发的程序。

3) 文[6]提出了使用 PePPeR(Perception Projection Performance Reliability)模型来设计性能测试;用户作为 Web 站点是否成功的最后的判定者,感知到的性能是很关键的,所以从用户的观点来设计性能测试。一般来说,用户希望所访问的站点的性能良好,所感知到的性能好的网站到底是什么样的,也说不清楚,所以把用户所感知到的网站的性能如响应时

间,吞吐量等进行系统地分析,设计成可以量化的指标,建立性能模型预测站点是否能够满足用户感知到的性能,最后设计额外的性能测试以保证站点的可靠性。这种方法对用户行为进行建模的方式设计性能测试,可以增强用户对网站的性能的信心。

4) 文[19]提出了一种测量 Web 站点性能的模型,主要是利用 CapCal™ ULTRA(Universal Load Testing Replay Agent),产生大量的虚拟用户来测试 Web 在可变的负载下的性能。传统方法主要是利用专门的机器来生成虚拟负载的,而 CapCal™ ULTRAs 主要是使用在网络上分布的代理来增加虚拟用户的数目,每一个代理可以模拟大量用户和会话,使用主机的 CPU 空闲时间生成负载。CapCal™ 结构是分布式的,利用了分布式计算来处理 Web 负载的动态性和不可预测性不需要专门的硬件来产生负载,并且利用了因特网以及公司安装的全部的计算机容量来做性能测试,这样能够尽可能地模拟真实的情况,并且降低了费用。

5) 商业性能测试工具主要有 Mercury LoadRunner, Radview WebLoad, Seague SilkTest, Rational Performance Tester 等。LoadRunner^[7],是一种预测系统行为和性能的工业标准级负载测试工具。使用有限的硬件资源,通过以模拟上千万用户实施并发负载及实时性能检测的方式来确认和查找问题,LoadRunner 能够对整个企业架构进行测试。通过使用 LoadRunner,企业能够最大限度地缩短测试时间,优化性能和加速应用系统的发布周期。LoadRunner 能支持广泛的协议和技术。Radview WebLoad^[8]性能测试工具可以让 Web 应用程序开发者通过模拟真实用户的操作,生成大量压力负载来测试 Web 的性能。WebLoad 提供巡航控制器 cruise control 的功能,可以预定义 Web 应用程序应该满足的性能指标,然后测试系统是否能满足这些指标,可以设定 WebLoad 采用自增用户数的循环方式进行测试,这样可以自动测试得到系统的最大用户容量。

6) 免费工具 WAS(Web Application Stress Tool)^[9],专门用来进行实际网站压力测试的一套工具,用来模拟 Web 浏览器对使用 http1.0 或 1.1 标准的 Web 服务器的请求,而不用考虑 Web 服务器运行在哪种平台上。使用 WAS,可以用不同的方式产生测试脚本。与其他工具不同的地方在于,WAS 可以使用多个客户端机器测试 Web 站点,把其中一个客户端作为主客户端用于协调其他客户端的测试。

7) 开放源性能测试工具 OpenSTA(Open, Systems Testing Architecture)^[10],是用 C++ 语言开发的软件,可以执行分布式测试,通过简单的图表形式和分布的测试,对于 HTTP 测试提供了很好的性能,对于简单的和可靠的 HTTP 测试来说是很好的软件。

国内在 Web 性能测试方面的研究和开发虽然刚刚开始,已经在这方面有了一些工作,文[11]中根据 Web 应用的特点,按应用端,中间网络和服务端三个方面对性能测试内容进行总结,抽象出了三层结构的性能测试模型,对性能测试的指标进行了总结,主要是对性能测试方面进行理论性的研究;文[12]开发了一个新的 Web 性能测试工具 Webmark,采用事件驱动的方法管理异步 I/O,通过修改 Client 的 TCP/IP 协议栈来模拟 Internet 环境,采用可靠的方法,模拟真实负载。文[13]提出了一种应用 TTCN(Tree Tabular Combine notion)形式化描述的 WAE(WEB Application Environment)性能测试方法,实现了一个基于 TTCN3 技术的 WAE 性能测试原型,即 VWPT(Visual WAE Performance TestWare)。梁晟等^[14]提出了模拟驱动的自动负载测试方法,并对 Web 应用

程序运行响应时间进行了研究,发现 Web 事务的响应时间和用户数量之间呈线性关系,测试的时间和测试配置对响应时间也有一定的影响。马琳等^[15]提出了一种基于转导推理的预测算法,目的主要是解决在测试资源有限的情况下,估计系统运行的可能行为和性能表现。

5 基于真实性能测试需要考虑的一些性能指标

1) 响应时间 响应时间也称为用户的等待时间,从用户的角度来看,指的是从用户发出请求至收到服务器的响应所经历的延迟。一般来说,使用站点的用户数越少,服务器处理的请求也就越少,所以响应时间就会比较短,当用户数目增多时,服务器比较繁忙,也就是说服务器要承受的负载会比较多,响应时间就会增加。

当 Web 开始出现的时候,人们使用网络的时候还觉得非常的新颖,他们对站点的性能有着很强的忍受程度。后来,随着人们对 Web 越来越熟悉,Internet 的推广成熟,基于 Web 的应用越来越普遍,人们只需要使用简单的方法,就可以很迅速方便地取得丰富的信息资料。网站,作为 Internet 中最基本的信息节点,为用户提供各种信息。如何迅速响应用户的访问成为了优秀网站的一个重要标志。对于商业网站来说更为重要。响应时间长的 Web 应用会使用户放弃其站点,而转向其他的响应时间快的站点,这样对于在网上进行其商品销售的厂家来说,无疑会减少潜在的客户,会造成很大的损失。Zona Research 的一项研究表明^[16],每年由于不能接受的下载速度导致的损失达十几亿美元。

Web 响应时间与很多因素有关,比如 HTTP 协议,Web 服务器,网络带宽,用户的位置等等。在测试时要考虑在真实环境中的响应时间,以便得到真实的负载。文^[18]第一次尝试在一个真实的环境中,对响应时间进行建模,通过监视真实的网站的响应时间行为,提出了一些有意义的结论。

2) 吞吐量 吞吐量是指在某一个特定的时间单位内 Web 应用所处理的用户请求数目。总的来说,用户的请求数目越多,吞吐量就会越大,但是当用户请求达到 Web 所能够并发处理的最大用户请求数目的时候,此时 Web 应用所处理的用户请求数目就是最大的吞吐量。通过考虑指标吞吐量可以找到系统的瓶颈问题。

3) 用户放弃率 用户作为 Web 站点成功与否的最终的决定者^[6],他感知的性能是至关重要的,不过由于众多的用户需求太复杂了,所以很难满足,最直观的反应就是站点的响应时间。响应时间的长短是判断用户满意程度的最重要的因素之一。到底响应时间达到什么程度才能保证站点有较好的性能呢,这就要使用互联网市场再三引用的“8 秒钟准则”,即当用户下载一个页面的时间超过 8 秒的时候,他们可能会放弃当前的网站而转移到别的网站来满足他们的要求。所以,用户放弃率在性能测试时也要考虑进去,否则,创建的负载测试在实际情况中永远不可能发生,找到的瓶颈也是不存在的,最重要的是在测试过程中也会忽略由于性能问题而放弃站点的用户数目,也就是说,我们创建的负载测试不是真实的。

4) 用户行为 在开放的网络中,用户行为是不可预见的,也就是说当用户在使用网络的时候,他们的行为是不同的,有些用户的点击速度要比其他用户的快一些,这样在特定的时间内,这些速度较快的用户将会产生更多的负载。一般来说,对网站比较熟悉的用户浏览一个页面用的时间比较短,反之用的时间要相对较长一些,因为他要有一个熟悉网站的过程,知道下一步该干什么。对于自己比较感兴趣的网站,不同用户对响应时间的忍受程度也是不一样的,在负载测试过程中考虑用户对网站的熟悉程度以及对网站的兴趣的大小将有助

于设计出精确的测试。

用户在使用网络的时候可能也在做其他的事情,比如在浏览电子邮件的时候下载其他的资料或者在吃饭的同时浏览新闻等等。在面在下载速度很慢的时候也有不同的操作,有的用户可能会在等待一个网站下载的时候转到另一个网站去找自己感兴趣的东西,而有的用户则会转去做其他的事情,而当时的下载就被放置在一边不去管了,还有的用户就不再等待而放弃了原来的操作。这些在实际情况中都会发生的,所以在设计负载测试时要尽可能地考虑这些因素。

在 Web 性能测试过程中,需要考虑的性能测试指标还有很多,例如页面请求次数、用户会话、页面请求分布等等,在性能测试过程中要考虑全面的性能测试指标。

总结及对未来的展望 Web 性能测试能够暴露出系统的性能问题,并提供相应的数据来帮助诊断和查明问题的原因。在性能测试过程中要尽量设计能够反映真实情况的性能测试,现有的方法和工具都以这个原则作为前提条件,因为不准确的性能测试容易对企业产生误导作用,测试结果也不会对 Web 的性能做出准确地指导。这也是现在性能测试方法和过程中要努力实现的目标。

为了让性能测试结果更加有说服力,要提供定量的分析让测试结果更清楚的表现出来,就要在测试过程中考虑更全面的性能测试指标,由于 Web 是不确定的和动态发展的,性能测试的指标在以后的发展过程中也要更新和补充。

Web 应用设计的复杂性及其具有分布、异构、并发和平台无关的特性,使得 Web 性能测试要比一般的性能测试复杂得多。对于 Web 性能测试,存在的难点主要有负载的不可预知性,测试场景设计的困难性,测试环境和真实环境的差异性^[14]。在这种情况下,一种普遍的方法就是在真实的环境中进行测试,测试的结果也是真实可信的,才会适用于测试已经发布的应用程序。为了得到真实可信的结果,要模拟真实的环境,可以运用形式化^[13]的方法建立规范完善的测试框架;又由于测试工作量太大,可以使用预测算法^[15]根据已有的测试数据进行分析预测。

参考文献

- Zona Research Inc. The Need for Speed. July 1999
- Subraya B M, Subraya S V. Design for Performance using PePPeR Model. <http://www.softwaredioxide.com/Channels/Content/Infosys-Design-Performance-PeppeR.pdf>
- Mercury Interactive Corp. Load testing to predict Web performance. 2000. http://wp.bitpipe.com/resource/org-968339186-562/WP-1079-0604-load_testing-bitpipe.pdf
- Savoia A. The science of website load testing. 2000. <http://www.keynote.com/downloads/whitepapers/science-of-loadtesting.pdf>
- Savoia A. The Science and Art of Web Site Load Testing. 2000
- Subraya B M, Subraya S V. Design for Performance using PePPeR Model. <http://www.softwaredioxide.com/Channels/Content/Infosys-Design-Performance-PeppeR.pdf>
- 刘艳会. LoadRunner 使用手册. 软件测试中心
- <http://www.radview.com/products/WebLOAD.asp>
- <http://www.west-wind.com/presentations/webstress/webstress.htm>
- <http://opensta.org/>
- 袁才国. Web 性能测试研究及工具开发: [学位论文]. 西北工大, 2004
- 张广艳. Web 性能测试: [硕士学位论文]. 吉林大学, 2003
- 张卫星. 基于 TTCN 的 WAE 性能测试研究: [学位论文]. 中国科学技术大学, 2002
- 梁晟, 李明树, 梁金能, 等. 一种模拟驱动的 Web 应用程序性能测试方法, 计算机研究与发展, 2003, 40(7)
- 马琳, 罗铁坚, 宋进亮, 等. Web 性能测试与预测. 中国科学院研究生院学报, 2005, 22(7)
- Savoia A. Web Page Response Time. Software Testing and Engineering Magazine, 2001
- Sevcik N F-P J. Understanding How Users View Application Performance. July 2002
- Turrini E. Analyzing web response time. <http://www.newcastle.research.ec.org/cabernet/workshops/plenary/5th-plenary-papers/analysingWebPerformance.pdf>
- Distributed Computing, Inc. A New Model For Measuring Web Site Performance. www.CapCal.com