

性能测试设计与执行

软件质量保障与测试课程 Lab6 课程作业（第 9 组）

Tian, Jiahe^{*} Hu, Xiaoxiao[†] Huang, Jiani[‡] Liu, Jiaxing[§]
Shi, Ruixin[¶] Wu, Chenning^{||} Zhang, Cenyuan^{**}
Zhang, Yihan^{††} Wang, Chen^{‡‡}

2020 年 5 月 14 日

^{*}Equal Contribution, Fudan University, 17307130313 (tianjh17@fudan.edu.cn)

[†]Equal Contribution, Fudan University, 17302010077 (xxhu17@fudan.edu.cn)

[‡]Equal Contribution, Fudan University, 17302010063 (huangjn17@fudan.edu.cn)

[§]Equal Contribution, Fudan University, 17302010049 (jiaxingliu17@fudan.edu.cn)

[¶]Equal Contribution, Fudan University, 17302010065 (rxshi17@fudan.edu.cn)

^{||}Equal Contribution, Fudan University, 17302010066 (cnwu17@fudan.edu.cn)

^{**}Equal Contribution, Fudan University, 17302010068 (cenyuanzhang17@fudan.edu.cn)

^{††}Equal Contribution, Fudan University, 17302010076 (zhangyihan17@fudan.edu.cn)

^{‡‡}Equal Contribution, Fudan University, 16307110064 (wangc16@fudan.edu.cn)

性能测试设计与执行

软件质量保障与测试课程 *Lab6* 课程作业

摘要

本次作业为软件质量保障与测试课程的 Lab6 课程作业，需要我们以小组为单位完成对出题系统的性能测试。本文档分为两小节。第一小节介绍了本小组进行性能测试采用的策略；第二小节介绍了性能测试的结果及系统性能分析。

关键词

系统与软件工程; 系统与软件质量要求和评价; 测试文档

目 录	3
-----	---

目 录

摘要	2
关键词	2
1 测试策略	4
2 系统性能及测试结果分析	4
2.1 登录	4
2.2 创建考题	5
参考文献	7

1 测试策略

本次测试的对象是出题系统中的登录和创建考题功能。登录功能分为认证以及选择项目两部分，创建考题功能则包括主持人创建新考题，完成属性配置、角色分配等活动，到将考题保存为止。性能指标为系统同时在线 100 人，20 个并发访问。为了产生必要的负载，需要依赖工具来进行性能测试。鉴于作为测试对象的出题系统是一个 web 应用，所以选择了开源的测试工具 Jmeter 来进行测试。通过 Jmeter 产生的聚合报告中的响应时间、吞吐量等来对系统性能进行分析。测试中采用的策略包括并发测试和负载测试。并发测试主要用于检验系统是否能够达到给定的性能指标，即在给定条件下相应用户输入的能力是否达到要求。负载测试中则是通过不同并发数下的测试，对比响应时间、吞吐量等的变化，来分析系统性能的可扩展性，系统处理能力何时达到饱和状态，以及观察在并发数超出给定指标的情况下，系统能否继续正常运行。并发测试中，对应在线 100 人的要求的是发出 100 个登录请求；对应 20 个并发访问的要求，发出 20 个创建考题的请求。我们使用了 Jmeter 工具中的定时器来让请求同时发送，以观察并发状况下服务的行为表现。同一时间对后端服务进行调用能更好地发现资源竞争、资源死锁等问题。负载测试中，则是调整了并发数，将不同并发数下的聚合报告。对于在线人数的要求，进行了从 20 到 140 个用户同时在线的测试，每相隔 20 数量进行一次测试。对于并发访问的要求，进行了创建 20 到 100 个考题的测试，每次增量为 20。通过这一策略，逐步增加系统的负载，直到超出指标，来寻找系统的性能上限，系统的处理能力，及系统在高负载情况下的稳定性。

2 系统性能及测试结果分析

2.1 登录

1. 100 个用户同时在线

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
登录-认证	100	9404	9441	16812	17921	18779	1284	19297	0.00%	5.2/sec	30.21	4.04
登录-选择项目	100	1128	1077	1519	1664	2016	575	2601	0.00%	5.4/sec	83.32	7.73
TOTAL	200	5266	1554	15251	16812	18738	575	19297	0.00%	10.0/sec	107.19	11.14

从结果中可以看出，在并发数为 100 的情况下，认证步骤的平均响应时间为 9.4 秒，选择项目的平均响应时间为 1.1 秒，整个登录步骤的平均响应时间约为 10.5 秒。根据用户满意度曲线来看，在并发数 100 的情况下，系统的响应时间有些过长，性能指标没有很好的达到。

2. 20 ~ 140 个用户同时在线，每次递增 20 个用户，对系统性能进行并发测试

aggregate_login												
Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received KB/sec	Sent KB/sec
登录-认证	20	2204	1839	3743	3748	3990	1245	3990	0.000%	4.60405	26.9	3.6
登录-选择	20	989	1005	1200	1263	1336	572	1336	0.000%	5.43183	84.21	7.81
TOTAL	40	1597	1263	2755	3743	3990	572	3990	0.000%	7.98085	85.17	8.85
登录-认证	40	3927	4146	6258	6725	6981	1336	6981	0.000%	5.18874	30.31	4.05
登录-选择	40	1038	1046	1297	1361	1385	601	1385	0.000%	5.75705	89.25	8.28
TOTAL	80	2483	1381	5260	6258	6884	601	6981	0.000%	9.6258	102.73	10.68
登录-认证	60	6773	6983	11221	11529	11854	1497	12051	0.000%	4.75473	27.78	3.71
登录-选择	60	1238	1046	2342	2534	2582	573	3200	0.000%	5.11771	79.34	7.36
TOTAL	120	4005	2101	9981	11221	11854	573	12051	0.000%	9.07578	96.86	10.07
登录-认证	80	7684	7167	13784	14484	14988	1317	15171	0.000%	5.22398	30.52	4.08
登录-选择	80	1163	1075	1525	1829	2274	697	2616	0.000%	5.43036	84.19	7.81
TOTAL	160	4424	1527	12627	13784	14940	697	15171	0.000%	9.94345	106.12	11.03
登录-认证	100	9404	9441	16812	17921	18779	1284	19291	0.000%	5.17117	30.21	4.04
登录-选择	100	1128	1077	1519	1664	2016	575	2601	0.000%	5.37461	83.32	7.73
TOTAL	200	5266	1554	15251	16812	18738	575	19291	0.000%	10.04369	107.19	11.14
登录-认证	120	11529	11586	20718	21875	22914	1320	23111	0.000%	5.16929	30.2	4.04
登录-选择	120	1226	1128	1716	2203	2625	560	2673	0.000%	5.34783	82.91	7.69
TOTAL	240	6378	1761	18528	20718	22825	560	23111	0.000%	10.09506	107.74	11.2
登录-认证	140	13185	12746	24200	25343	26578	1270	26761	0.000%	5.18135	30.27	4.05
登录-选择	140	1182	1172	1547	1628	1821	611	1936	0.000%	5.31027	82.32	7.63
TOTAL	280	7184	1607	21249	24200	26425	611	26761	0.000%	10.13245	108.14	11.24

在上一小节中，我们小组根据助教给出的指标进行了测试，发现测试的情况并不能达到在 100 并发的情况下有较好的响应性能。因此，我们小组进一步进行了更为深入的并发测试。在这一测试中，我们选取了上一小节中测试的并发数的 20% 作为初始值，并以上述并发数的 20% 作为步长进行阶梯并发测试，最终以上述并发数目的 140% 作为终止值。从而进一步探究这一系统在不同的并发数下的性能情况，并从这一测试的结果中得到这一系统能够支撑的较大并发数目。

下面对这一部分测试的结果情况进行分析：

从表中可以看出，登录的选择部分的响应时间随着并发数的增加变化不大，说明这一服务有较大的可扩展性。而认证部分随着并发数增加，响应时间明显变长。所以对于登录功能来说，认证部分是性能的瓶颈所在，调优时应重点关注。

并发数达到 80 后，吞吐量不再有太大的增长，说明系统处理能力已经接近饱和。

登录部分的性能测试未能达到指标，可能是由于在本地服务器上进行的测试。但在并发数超过指标 20% 以上的情况下，虽然响应时间变长，但系统仍然能保持功能完整性，没有失效，说明系统在压力下仍能正常运行，稳定性较好。

2.2 创建考题

1. 20 道考题的并发创建

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received KB/sec	Sent KB/sec
新建题目	20	1931	2090	2522	2523	2557	931	2557	0.00%	7.8/sec	32.35	18.13
TOTAL	20	1931	2090	2522	2523	2557	931	2557	0.00%	7.8/sec	32.35	18.13

从结果中可以看出，在并发数为 20 的情况下，响应时间平均为 1.9 秒，最大不超过 2.5 秒。所以创建考题部分较好地达到了性能指标。

2. 20 ~ 100 道考题并发创建，每次递增 20 道考题创建量，对系统性能进行并发测试

aggregate_create_question

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Max	Error %	Throughput	Received KB/sec	Sent KB/sec
新建题目	20	1849	2034	2429	2459	2460	864	2460	0.000%	8.12678	33.63	18.84
TOTAL	20	1849	2034	2429	2459	2460	864	2460	0.000%	8.12678	33.63	18.84
新建题目	40	2758	2717	3607	3730	4249	1090	4249	0.000%	9.40734	38.92	21.81
TOTAL	40	2758	2717	3607	3730	4249	1090	4249	0.000%	9.40734	38.92	21.81
新建题目	60	4124	3797	6550	7009	7169	987	7334	0.000%	8.17439	33.82	18.95
TOTAL	60	4124	3797	6550	7009	7169	987	7334	0.000%	8.17439	33.82	18.95
新建题目	80	4243	3990	6818	7131	7202	1226	7329	0.000%	10.91107	45.15	25.3
TOTAL	80	4243	3990	6818	7131	7202	1226	7329	0.000%	10.91107	45.15	25.3
新建题目	100	6114	6116	8589	8849	9074	1056	9217	0.000%	10.84246	44.86	25.14
TOTAL	100	6114	6116	8589	8849	9074	1056	9217	0.000%	10.84246	44.86	25.14

从表中可以看出，随着并发数的上升，响应时间有明显的增加。但在超过指标 100% 以上的情况下，服务仍然能够正常运行，且以用户满意度曲线为参考，响应时间较为合理，所以这一服务拥有很高的可扩展性。吞吐量同样在并发数达到 80 后趋于稳定，说明系统处理能力趋于饱和。

参考文献

International Organization for Standardization. 2014. *Systems and Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE*. International Organization for Standardization. Vol. 2014. <https://www.iso.org/standard/64764.html>.

中国国家标准化管理委员会. 2016. *GB/T 25000.51-2016《系统与软件工程系统与软件质量要求和评价 (SQuaRE) 第 51 部分: 就绪可用软件产品 (RUSP) 的质量要求和测试细则》*. 系统与软件工程系统与软件质量要求和评价 (SQuaRE). Vol. 51. 中国国家标准化管理委员会. <http://openstd.samr.gov.cn>.

———. 2017a. *GB/T 25000.12-2017《系统与软件工程系统与软件质量要求和评价 (SQuaRE) 第 12 部分: 数据质量模型》*. 系统与软件工程系统与软件质量要求和评价 (SQuaRE). Vol. 12. 中国国家标准化管理委员会. <http://openstd.samr.gov.cn>.

———. 2017b. *GB/T 25000.24-2017《系统与软件工程系统与软件质量要求和评价 (SQuaRE) 第 24 部分: 数据质量测量》*. 系统与软件工程系统与软件质量要求和评价 (SQuaRE). Vol. 24. 中国国家标准化管理委员会. <http://openstd.samr.gov.cn>.

———. 2018. *GB/T 25000.40-201《系统与软件工程系统与软件质量要求和评价 (SQuaRE) 第 40 部分: 评价过程》*. 系统与软件工程系统与软件质量要求和评价 (SQuaRE). Vol. 40. 中国国家标准化管理委员会. <http://openstd.samr.gov.cn>.

———. 2019. *GB/T 25000.23-2019《系统与软件工程系统与软件质量要求和评价 (SQuaRE) 第 23 部分: 系统与软件产品质量测量》*. 系统与软件工程系统与软件质量要求和评价 (SQuaRE). Vol. 23. 中国国家标准化管理委员会. <http://openstd.samr.gov.cn>.