分布式爬虫

测试报告

班级 信1405-2班

姓名 邵文政 陈鹏 祖浩然

目录

**[1 概述](#_Toc326412934)** [3](#_Toc326412934)

[1.1 目的 3](#_Toc326412935)

[1.2 背景 3](#_Toc326412936)

[1.3 范围 3](#_Toc326412937)

**[2 测试概要](#_Toc326412938)** [3](#_Toc326412938)

[2.1 测试环境 3](#_Toc326412939)

[2.2 人力资源 3](#_Toc326412940)

[2.3 测试工作量 4](#_Toc326412941)

**[3 测试内容及方法](#_Toc326412942)** [4](#_Toc326412942)

[3.1 测试需求/目标 4](#_Toc326412943)

[3.2 测试内容 4](#_Toc326412944)

[3.3 测试工具 4](#_Toc326412945)

**[4 测试结果及分析](#_Toc326412946)** [4](#_Toc326412946)

[4.1 网站处理性能评估 4](#_Toc326412947)

[4.2并发登录用户测试 5](#_Toc326412948)

**[5 结果分析](#_Toc326412949)** [6](#_Toc326412949)

[5.1 场景执行情况 6](#_Toc326412950)

[5.2 Statistics Summary（统计信息摘要） 6](#_Toc326412951)

[5.3 Transaction Summary（事务摘要） 7](#_Toc326412952)

[5.4 HTTP Responses Summary（HTTP响应摘要） 7](#_Toc326412953)

[5.5 并发数分析 8](#_Toc326412954)

[5.6 响应时间 9](#_Toc326412955)

[5.7 每秒点击数 9](#_Toc326412956)

# 1 概述

## 1.1 目的

本测试报告爬虫性能测试报告，目的在于总结测试阶段的测试以及分析测试结果，描述网站是否符合需求。

## 1.2 背景

考虑到用户数量及数据的增多给服务器造成压力不可估计，因此计划对网站负载性能测试，在系统配置不变的情况下，在一定时间内，服务器在高负载情况下的性能行为表现，便于对系统环境进行正确的分析及评估。

## 1.3 范围

本次测试主要是分布式爬虫性能测试。

# 2 测试概要

## 2.1 测试环境

PC机：Lenovo笔记本

操作系统：windows10

测试机与被测服务器在同一局域网进行，排除了网速限制及网速度不稳定性。

## 2.2 人力资源

下表列出了所有参与此项目的测试人员：

|  |  |
| --- | --- |
| 角色 | 资源数量/具体人员 |
| 测试员 | 祖浩然 |

## 2.3 测试工作量

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 任务 | 开始时间 | 结束时间 | 总计（天数） | 总计（人时） |
| 计划 | 2012/6/2 | 2012/6/5 | 3 | 3人·天 |
| 实际 | 2012/6/5 | 2012/6/6 | 1 | 1人·天 |

# 3 测试内容及方法

## 3.1 测试需求/目标

在大用户量、数据量的超负荷下，获得服务器运行时的相关数据，从而进行分析系统的稳定性。

## 3.2 测试内容

本次测试主要是对百度首页访问操作在大负荷情况下处理数据的能力及承受能力。

测试方法：

|  |  |  |  |
| --- | --- | --- | --- |
| 场景 | 并发用户数量 | 运行场景设置 | 测试点 |
| 登录 | 200 | 2分钟 | 服务器稳定性及操作响应时间 |

## 3.3 测试工具

主要测试工具为：LoadRunner性能测试工具

辅助软件：谷歌浏览器 Word2007

# 4 测试结果及分析

## 4.1 网站处理性能评估

这次测试属于局域网环境进行，排除了外网的网速限制及不稳定性。

## 4.2并发登录用户测试

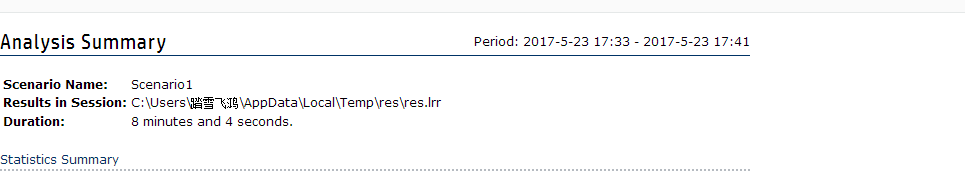
测试内容：

说明：用户的整个执行流程都录制在Action(循环)部分，所以Vuser\_int (开始)和Vuser\_end(结束)部分为空。Action\_Transaction部分的时间为运行整个Action脚本所需的时间。

整个Action的平均响应时间为：71.77秒。

# 5 结果分析

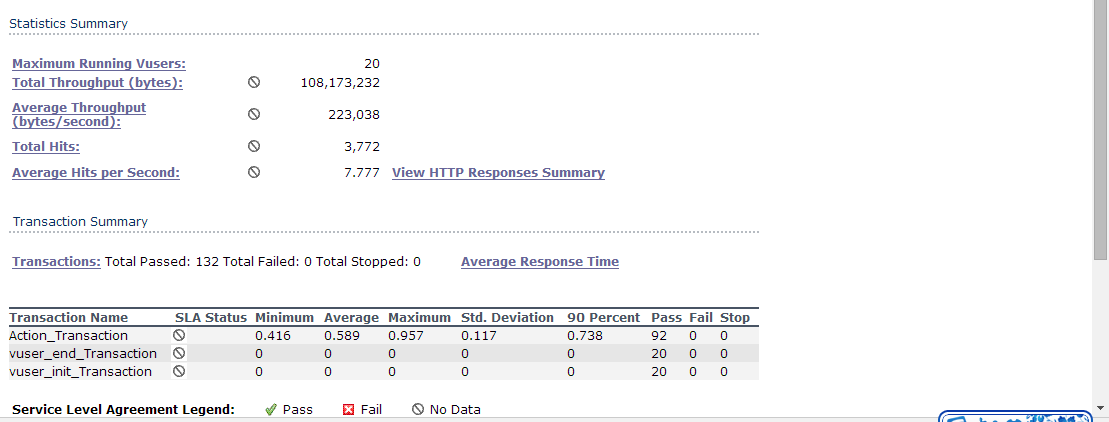
## 5.1 场景执行情况





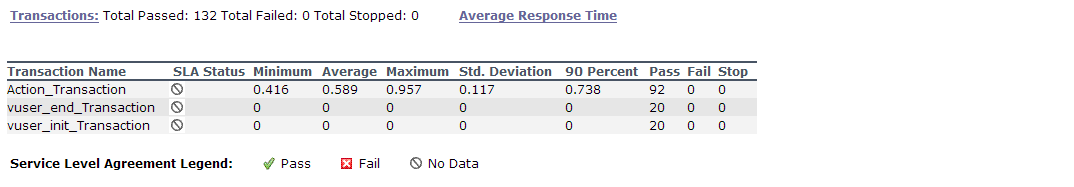
该部分给出了本次测试场景的名称、结果存放路径及场景的持续时间，如上图所示。从该图我们知道，本次测试从2017/5/23 17:33 开始，到2017/5/23 17:41结束，共历时8分。

## 5.2 Statistics Summary（统计信息摘要）



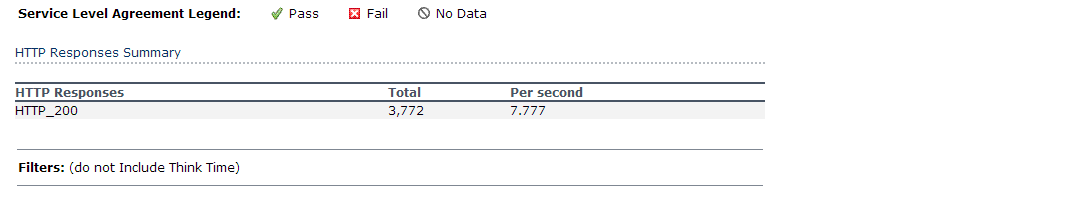
该部分给出了场景执行结束后并发数、总吞吐量、平均每秒吞吐量、总请求数、平均每秒请求数的统计值，如图所示。从该图我们得知，本次测试运行的最大并发数为20，总吞吐量为108,173,232字节，平均每秒的吞吐量为223,038字节，总的请求数为3,772，平均每秒的请求为7.777，对于吞吐量，单位时间内吞吐量越大，说明服务器的处理能越好，而请求数仅表示客户端向服务器发出的请求数，与吞吐量一般是成正比关系。

## 5.3 Transaction Summary（事务摘要）



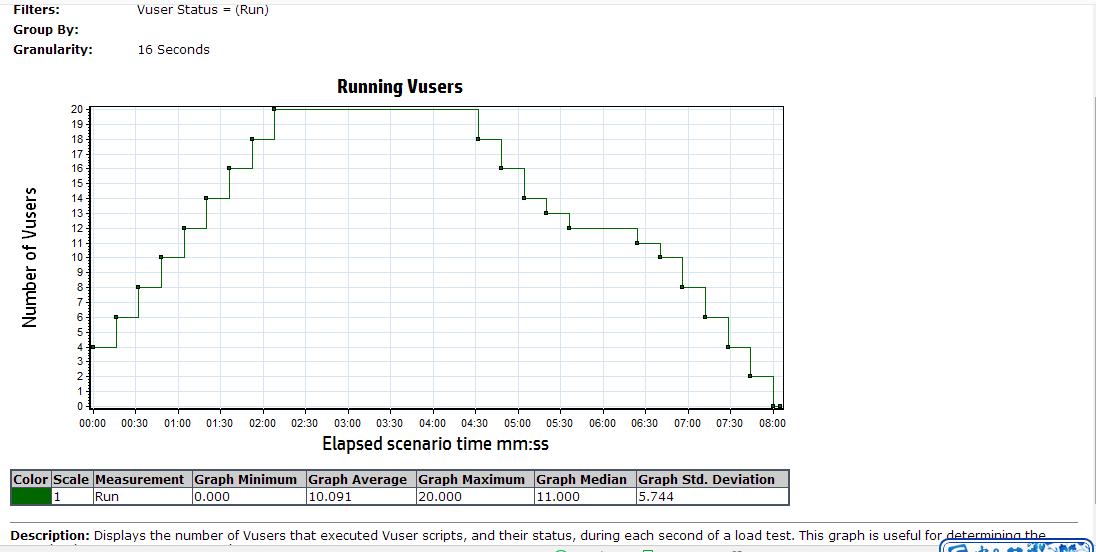
该部分给出了场景执行结束后相关Action的平均响应时间、通过率等情况，如上图所示。从该图我们得到每个Action的平均响应时间与业务成功率。

## 5.4 HTTP Responses Summary（HTTP响应摘要）



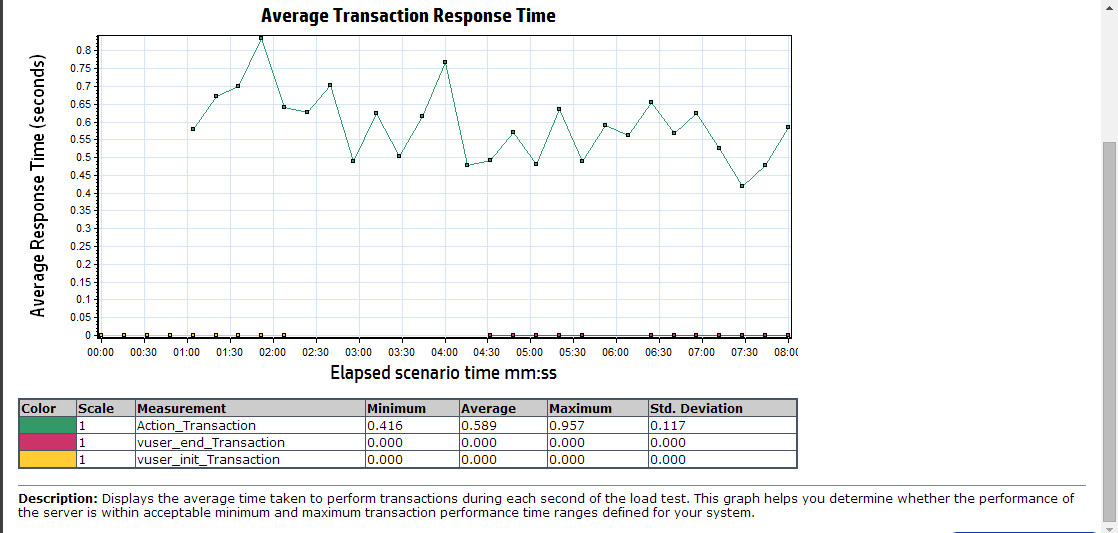
该部分显示在场景执行过程中，每次HTTP请求发出去的状态，是成功还是失败，都在这里体现，如图5- 6所示。从图中可以看到，在本次测试过程中LoadRunner共模拟发出了3772次请求（与“统计信息摘要”中的“Total Hits”一致），其中“HTTP 200”的是6190次，而“HTTP 204”则有23，说明在本次过程中，经过发出的请求大部分都能正确响应了，但还是有部分未得到任何返回内容，但未影响测试结果，“HTTP 200”表示请求被正确响应，而“HTTP 204”表示服务器成功处理了请求，但未返回任何内容。

## 5.5 并发数分析



“Running Vusers（运行的并发数）”显示了在场景执行过程中并发数的执行情况。它们显示Vuser的状态、完成脚本的Vuser的数量以及集合统计信息，将这些图与事务图结合使用可以确定Vuser的数量对事务响应时间产生的影响。上图显示了百度性能测试过程中Vusers运行情况，从图中我们可以看到，Vusers的运行趋势与我们场景执行计划中的设置是一样，表明在场景执行过程中，Vusers是按照我们预期的设置运行的，没有Vuser出现运行错误，这样从另一个侧面说明我们的参数化设置是正确的，因为使用唯一数进行参数化设置，如果设置不正确，将会导致Vuser运行错误。

## 5.6 响应时间

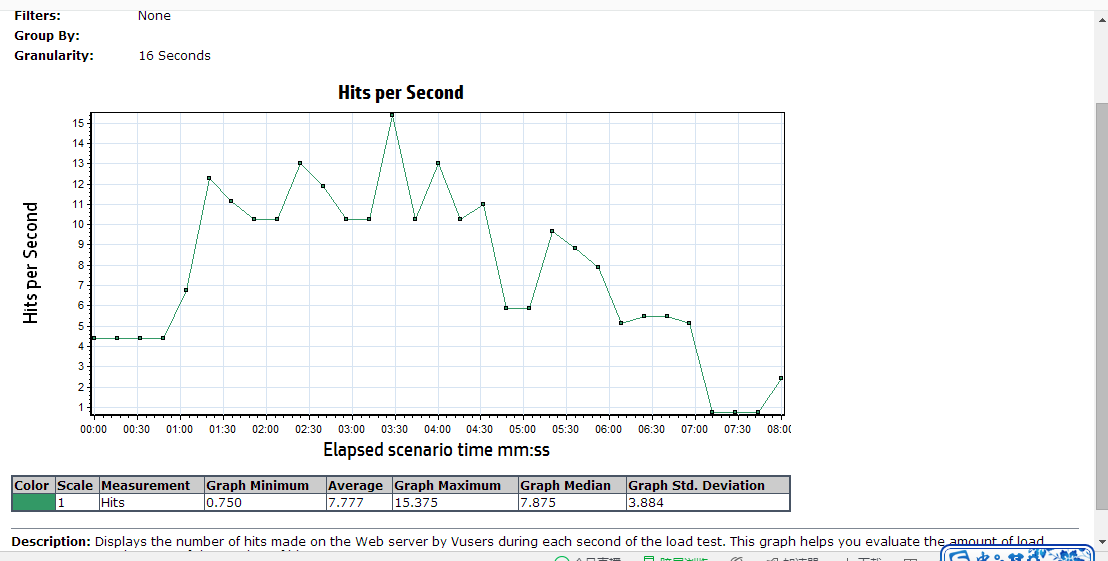


这张图是平均事务响应时间与结果摘要中的“Transaction Summary”合成的。

此次测试用户操作流程简单，但20并发用户对服务器造成高度负载，服务器运行不稳定。

从设置20人的压力分析，响应速度太慢，超出用户的感觉快速响应时间。

## 5.7 每秒点击数



“Hits per Second（每秒点击数）”反映了客户端每秒钟向服务器端提交的请求数量，如果客户端发出的请求数量越多，与之相对的“Average Throughput (bytes/second)”也应该越大，并且发出的请求越多会对平均事务响应时间造成影响，所以在测试过程中往往将这三者结合起来分析。从图中可以看出， “Hits per Second”正常，而“Average Throughput (bytes/second)”不正常，则表示服务器虽然能够接受服务器的请求，但返回结果较慢，可能是程序处理缓慢。