

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра материаловедения, технологии
и управления качеством

**УПРАВЛЕНИЕ КАЧЕСТВОМ ДАННЫХ ПРИ ТЕСТИРОВАНИИ
ПРОИЗВОДИТЕЛЬНОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента магистратуры 2 курса 2301 группы
направления 27.04.02 «Управление качеством»,
профиль «Менеджмент качества в инженерной и образовательной
деятельности»
института физики

Сагитова Романа Павловича

Научный руководитель,
доцент, к.ф.-м.н., доцент

должность, уч. степень, уч. звание

подпись, дата

Д.В. Терин

инициалы, фамилия

Зав. кафедрой,
д.ф.-м.н., профессор

должность, уч. степень, уч. звание

подпись, дата

С.Б. Вениг

инициалы, фамилия

Саратов 2024

Введение. Современные наукоемкие технологии все чаще работают с так называемыми Большими Данными. Высокий темп роста в сегменте систем хранения больших данных связан с особенностями их использования в настоящий момент.

На сегодняшний день большинство крупных компаний испытывает трудности, связанные со стремительным ростом общего объема информации. По всем прогнозам, скорость прироста будет все более ускоряться, а это значит, что большие данные требуют новых, более современных подходов к их хранению и управлению.

Управление качеством в области Больших Данных должно быть реализовано согласно общим стандартам менеджмента качества и стандартам менеджмента качества для IT технологий.

В современном мире качество является ключевым фактором, определяющим успех любого предприятия. Организация ООО «ЛАНИТ Экспертиза» также не исключение. Эта компания, специализируется на обеспечении качества информационных систем, понимает важность качества и стремится к его постоянному улучшению. Компания предоставляет полный спектр услуг в области независимого тестирования информационных систем, включая тестирование производительности.

Целью данной работы является изучение процессов, функционирующих в отделе по тестированию производительности, анализ влияния качества данных на результаты тестирования и описание используемых решений в области управления качеством данных при тестировании производительности информационных систем.

Исходя из цели работы, выделим ключевые задачи, которые предстоит решить:

1. изучить основные термины и содержание управления качеством данных;
2. представить общую характеристику организации ООО «ЛАНИТ Экспертиза» (направления деятельности, виды оказываемых услуг, провести

анализ среды организации, выделить ее потребителей и заинтересованных сторон);

3. идентифицировать процессы, функционирующие в отделе по тестированию производительности;

4. идентифицировать риски процесса нагрузочного тестирования, предложить мероприятия (действия) по их минимизации и/или устранению;

5. проанализировать влияние качества данных на результаты тестирования производительности информационных систем;

6. описать используемые решения в области управления качеством данных при тестировании производительности информационных систем.

Выпускная квалификационная работа занимает 54 страницы, имеет 19 рисунков и 5 таблиц.

Обзор составлен по 22 информационным источникам.

Во введение рассматривается актуальность работы, устанавливается цель и выдвигаются задачи для достижения поставленной цели.

Первый раздел представляет собой теоретические основы управления качеством данных.

Во втором разделе работы производится обработка и анализ информации об организации ООО «ЛАНИТ Экспертиза», идентификация процессов, функционирующих в отделе по тестированию производительности, определение рисков процесса нагрузочного тестирования и предложение мероприятий по их минимизации и/или устранению.

В третьем разделе работы проводится исследование влияния качества данных на результаты нагрузочного тестирования, и описываются используемые решения в области управления качеством данных при тестировании производительности информационных систем

Основное содержание работы

В современном информационном ландшафте, насыщенном множеством данных, эффективное управление информацией становится неотъемлемой частью стратегического развития предприятий. В эпоху больших данных, где

объемы информации экспоненциально возрастают, необходимость систематизации, структурирования, и обеспечения качества данных становится жизненно важной. Ключевыми вызовами в данном контексте являются не только эффективная обработка и хранение данных, но и обеспечение их точности, актуальности и согласованности. Модернизация методов управления данными нацелена на реализацию автоматизированных и интеллектуальных систем, способных обнаруживать и устранять ошибки, обеспечивая высокий стандарт качества данных.

В 1968 году была предложена концепция базы данных, которая подразумевала дальнейшее развитие. В 1981 году Э. Фрэнк Кодд окончательно сформулировал реляционную модель данных, которая легла в основу разработки реляционных баз данных и программного стандарта управления данными, основанного на языке SEQUENCE, который позже был переименован в SQL. Реляционная модель данных и SQL по-прежнему очень популярны при проектировании и внедрении баз данных.

В теории реляционных баз данных установлены важные понятия о независимости физического уровня хранения данных, логической модели и интерфейса управления данными. Стандартизированные методы, разработанные на основе правил «нормальной формы», позволяют достичь требуемого уровня согласованности данных с помощью внутренних механизмов программного обеспечения системы управления базами данных (СУБД).

Качество данных – обобщенное понятие, отражающее степень пригодности информации для решения соответствующей задачи. Оценка данных как «качественные» или «некачественные», а также ее разные градации (плохая, хорошая, отличная) – имеют субъективный характер и должны рассматриваться непосредственно в контексте решаемой задачи.

Существует ряд критериев, которые применяются для оценки правильности, полноты, точности и надежности данных. Наиболее используемые из них перечислены ниже.

Полнота – достаточность объема, глубины и широты наборов данных. Неполнота приводит либо к невозможности анализа, либо к необходимости отталкиваться от некоторых предположений или допущений относительно пропущенной информации. Полнота может касаться как пропусков в атрибутах анализируемых объектов, например, не заполненные сведения в справочнике товаров, так и отсутствия части исследуемых данных, например, сведений за определенный период.

Релевантность – показатель того, насколько данные соответствуют целям и решаемым задачам. Например, сведения о продажах бумажных книг могут быть нерелевантными для рынка электронных, либо данные о предпочтениях клиентов в одной стране могут полностью отличаться от пользователей в другой.

Валидность – соответствие многочисленным атрибутам, связанных с элементом данных: тип, точность, формат, диапазоны допустимых значений и так далее. Например, в поле e-mail строка должна соответствовать стандарту написания электронных адресов.

Точность – детальность измерения и фиксации данных. В зависимости от особенностей процесса и целей анализа показатели необходимо фиксировать с точностью до дня, часа, минуты или секунды, либо вес товара измерять с точностью до тонны или грамма.

Своевременность – время после сбора данных, по прошествии которого они становятся доступными для анализа. Оно должно соответствовать скорости анализируемого процесса. Корректные, но устаревшие данные бесполезны для принятия оперативных решений.

Уникальность подразумевает, что ни один объект не существует в наборе данных более одного раза. Наличие дублей может приводить к несогласованности и противоречиям вследствие отсутствия единой версии правды.

Целостность – наличие корректных ссылок между данными и их соответствие установленным правилам и ограничениям. Ссылочная

целостность предполагает, что все ссылки из данных в одном столбце таблицы на данные в другом столбце той же или другой таблицы являются допустимыми, т.е. не будет ситуации, при которой запись в таблице продаж ссылается на покупателя, отсутствующего в справочнике клиентов.

Согласованность – соответствие данных друг другу и их логическая непротиворечивость. Например, соответствие пола человека его имени, а даты рождения – возрасту. Если данные не согласованы, это может указывать на ошибки или неточности в их сборе или обработке.

Надежность – возможность повторного получения одинаковых результатов. Если результаты измерения получаются разные в зависимости от условий, то снижается доверие к принятым на их основе решениям.

Поиск, обработка и анализ информации об исследуемой организации:

Наименование исследуемой организации: ООО «ЛАНИТ Экспертиза».

Основная цель деятельности компании: обеспечение качества программного обеспечения и информационных систем. Компания предоставляет полный спектр услуг в области тестирования программного обеспечения, включая функциональное тестирование, интеграционное тестирование, тестирование производительности и другие виды тестирования. Кроме того, «ЛАНИТ Экспертиза» занимается обучением специалистов в области автоматизированного и нагрузочного тестирования, а также созданием команд инженеров тестирования под потребности заказчика, которыми являются крупные банки, ведущие российские IT-компании и др. государственные заказчики [1].

Основные процессы отдела по тестированию производительности [2]:

- Анализ требований производительности.
- Планирование тестирования.
- Настройка тестового окружения.
- Проведение тестирования.
- Анализ результатов.
- Оптимизация и рекомендации.

- *Документирование и отчетность.*

Этапы нагрузочного тестирования [3]:

- Получение задачи и анализ тестируемой системы.
- Разработка методики нагрузочного тестирования (МНТ).
- Подготовка к тестированию.
- Проведение тестирования.
- Обработка артефактов и анализ результатов.
- Составление отчета.
- Закрытие задачи.

Рекомендации по улучшению функционирования процесса:

- Четкое определение целей тестирования.
- Тщательный анализ требований.
- Автоматизация тестирования.
- Реалистичные сценарии нагрузки.
- Мониторинг и анализ результатов в реальном времени.
- Регулярное обновление тестовых сценариев.
- Управление рисками.
- Обучение и обмен опытом.
- Непрерывное улучшение процесса.

Проведем исследование влияния качества данных на результаты нагрузочного тестирования.

Цель исследования.

Определение влияния объема данных в базе данных на производительность системы HelpDesk при выполнении нагрузочного.

Объект тестирования.

Сервис обслуживания клиентов ИС HelpDesk, осуществляющий автоматизацию приёма заявок и их управление внутренней службой поддержки.

Этапы исследования.

1. Провести тест с объемом базы данных сервиса в 20 гигабайт.
2. Наполнение базы данных сервиса тестовыми данными SQL-скриптом на 100 гигабайт.
3. Провести тест с объемом базы данных сервиса в 120 гигабайт.
4. Определить влияния объема данных в базе данных на производительность системы HelpDesk.

Заключение по результатам исследования.

На рисунке 1 представлены времена отклика операций по двум тестам [2].

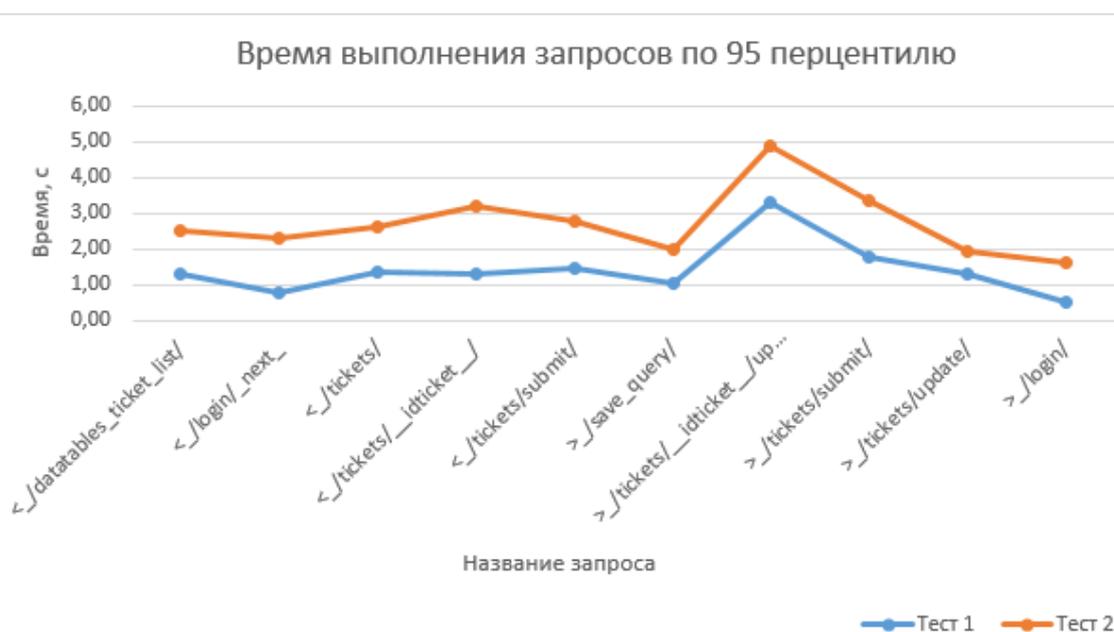


Рисунок 1 – Время выполнения запросов по 95 перцентилю

Результаты второго тестирования информационной системы привели к заметному увеличению времени отклика операций. В среднем времена выполнения запросов оказались выше на 1,3 секунды по сравнению с первым тестированием. Это явление может служить непосредственным свидетельством влияния объема данных в БД на производительность системы.

С учетом вышеописанных исследований, делаем вывод о важности поддержания высокого качества данных в процессе тестирования производительности. Предложение о следящих мерах включает в себя регулярный аудит данных, использование реалистичных сценариев тестирования, и внимательный мониторинг изменений в данных между различными этапами тестирования. Такие меры не только помогут

предотвратить искажение результатов тестирования, но и обеспечат более достоверную оценку производительности системы, что важно для принятия обоснованных решений по её оптимизации.

В каждой компании, которая проводит нагрузочное тестирование, имеется определённый инструмент или набор инструментов для обеспечения качества данных. Одним из таких инструментов является QAD – quality assurance dashboard. Он разработан в отделе по нагрузочному тестированию Хоум банка. В QAD много различного функционала, в интеграции с другими программными продуктами, такими как Jenkins, Kafka, стек ELK (Elasticsearch, Logstash и Kibana) и др., он образует единую систему централизации данных за определенный период времени с целью последующего анализа [4].

Можно выделить следующие типы анализа [5]:

- Оперативный анализ – анализ, при котором выдача результата для принятия решения о производительности системы по ряду показателей, которые могут рассчитываться и сравниваться в реальном времени.
- Приемочный анализ – анализ, при котором выдача результата для принятия решения о производительности системы осуществляется после сравнения всех собираемых метрик производительности из текущего теста с эталонным.
- Полный анализ – это расширение приемочного анализа посредством корреляции полученных показателей с исторически накопленными данными предыдущих успешных тестов.

Пример автоанализа в QAD представлен на рисунках 2, 3.

Список систем и статусы

СИСТЕМА	ОШИБКИ	ПВП	О	ВО	ЕК	НД	ИП
ANTIFRAUD	0 (0)	PASS	PASS	FAIL	PASS	PASS	PASS
HOMER	44 (42)	PASS	FAIL	PASS	PASS	PASS	PASS
HOMER-SAT	0 (0)				PASS	PASS	
IBS	0 (0)	PASS	PASS	PASS	PASS	PASS	PASS
MSRV	392 (436)	PASS	FAIL	PASS	FAIL	PASS	PASS
MYCREDIT	1679 (257)	PASS	FAIL	PASS	FAIL	FAIL	PASS
OSB	4 (1)	PASS	PASS	PASS	PASS	PASS	PASS
QUORUM	0 (0)	PASS	PASS	PASS	PASS	PASS	PASS
SMSGATE	0 (0)	PASS	PASS	PASS	PASS	FAIL	PASS
TWCMS	0 (0)	PASS	PASS	PASS	FAIL	PASS	PASS

Рисунок 2 – Приемочный автоанализ

На рисунке 2 представлен пример приемочного автоанализа в QAD. На данном этапе можно понять стоит ли рассматривать этот тест как успешный и проводить полный анализ.

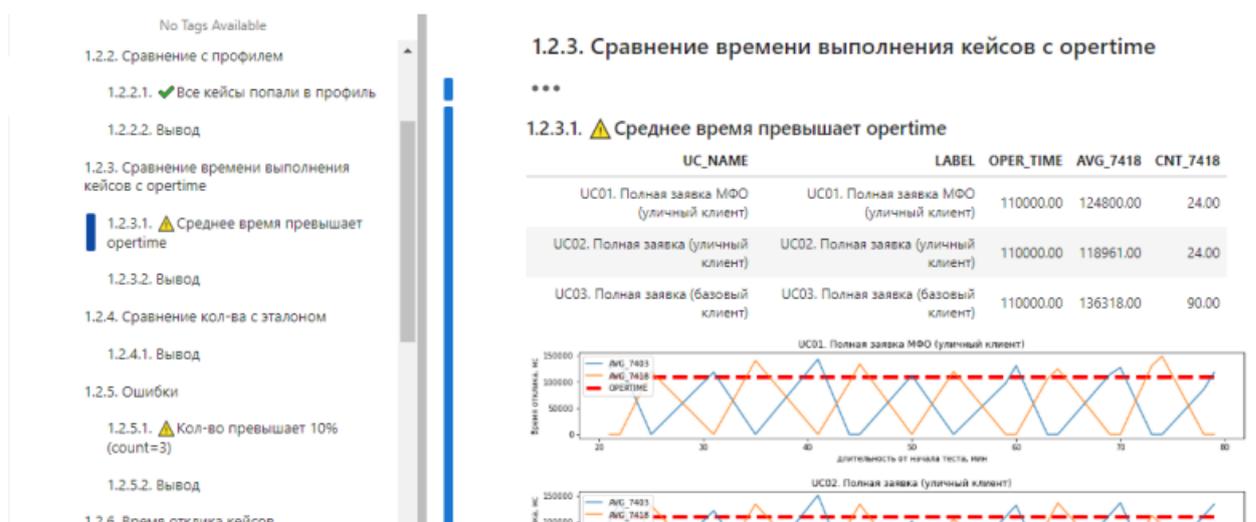


Рисунок 3 – Часть полного анализа из Jupyter Notebook

На рисунке 3 представлена часть полного анализа, выполненного в Jupyter Notebook с использованием языка программирования Python. На этом этапе специалист по нагрузочному тестированию проходит по всем пунктам, выделенным восклицательным знаком, и делает выводы о причинах отклонений. Результатом такого анализа является заключение о производительности информационной системы.

Заключение. В ходе данной работы были осуществлены ключевые задачи, направленные на изучение процессов, функционирующих в отделе по

тестированию производительности, анализ влияния качества данных на результаты тестирования и описание используемых решений в области управления качеством данных при тестировании производительности информационных систем.

Были изучены основные термины и содержание управления качеством данных, что предоставило основу для анализа влияния качества данных на производительность системы.

Представлена общая характеристика организации «ЛАНИТ Экспертиза», выделены направления деятельности, виды оказываемых услуг, проведен анализ среды организации, определены ее потребители и заинтересованные стороны.

Идентифицированы процессы отдела по тестированию производительности, что позволило понять их структуру и взаимосвязи.

Зафиксированы риски процесса нагрузочного тестирования, предложены действенные мероприятия по их минимизации и/или устранению, что способствует повышению эффективности тестирования.

Проанализировано влияние качества данных на результаты тестирования производительности информационных систем, подчеркнув необходимость внимательного контроля этого аспекта для достоверных результатов.

Рассмотрен функционал QAD – quality assurance dashboard. Который используется в области управления качеством данных при тестировании производительности информационных систем

Итоговые выводы подчеркивают важность системного подхода к управлению качеством данных и показывают, что эффективное нагрузочное тестирование требует не только технической компетентности, но и внимания к качеству данных, чтобы обеспечить точность и достоверность результатов.

Список использованных источников

1 Тестирование и экспертиза ИТ-продуктов [Электронный ресурс] // «ЛАНИТ Экспертиза» [Электронный ресурс] : [сайт]. – URL: <https://lanit-exp.ru/> (дата обращения: 07.12.2023). – Загл. с экрана. – Яз. рус.

2 Чехахин, А. В. Нагрузочное тестирование клиент-серверных приложений / А. В. Чехахин // Научный альманах Центрального Черноземья. – 2022. – № 1-1. – С. 116-127.

3 Туровец, Н. О. Цели и этапы проведения нагрузочного тестирования информационных систем / Н. О. Туровец // Электронные системы и технологии : материалы 59-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 17-21 апреля 2023 года. – Минск : Научное электронное издание, 2023. – С. 329-331.

4 Как мы автоматизировали и упростили процесс ведения релизов [Электронный ресурс] // Хабр [Электронный ресурс] : [сайт]. – URL: <https://habr.com/ru/companies/homebank/articles/799369> (дата обращения: 02.05.2024). – Загл. с экрана. – Яз. рус.

5 Сухов Денис. Интеллектуальный анализ – уменьшаем TTM [Электронный ресурс] : Youtube [Электронный ресурс] : [сайт]. – URL: <https://www.youtube.com/watch?v=yDJziFja7f4> (дата обращения: 02.05.24). – Загл. с экрана. – Яз. рус.