



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(национальный исследовательский университет)»

Институт (Филиал) № 8 «Компьютерные науки и прикладная математика» Кафедра 806
Группа М8О-406Б-21 Направление подготовки 01.03.02 «Прикладная математика и
информатика»

Профиль Информатика

Квалификация: бакалавр

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

на тему: Проектирование системы переноса и генерации взаимосвязанных
данных из производственной среды при тестировании образовательной
платформы

Автор ВКРБ:	Мезенин Олег Александрович	(_____)
Руководитель:	Миронов Евгений Сергеевич	(_____)
Консультант:	Ляпина Светлана Юрьевна	(_____)
Консультант:	—	(_____)
Рецензент:	—	(_____)

К защите допустить

Заведующий кафедрой № 806	Крылов Сергей Сергеевич	(_____)
---------------------------	-------------------------	---------

_____ мая 2025 года

Москва 2025

РЕФЕРАТ

Выпускная квалификационная работа бакалавра состоит из 18 страниц, 4 рисунков, 6 использованных источников, 1 приложения.

ПЕРЕНОС ДАННЫХ, РЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ, ТЕСТИРОВАНИЕ, АНОНИМИЗАЦИЯ ДАННЫХ, ГЕНЕРАЦИЯ ДАННЫХ, ГРАФЫ

Здесь могла бы быть ваша аннотация.

СОДЕРЖАНИЕ

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	4
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	5
ВВЕДЕНИЕ	6
1 Постановка задачи и теоретические предпосылки	9
1.1 Тестирование	9
1.1.1 Процесс тестирования	9
1.2 Ручное и автоматизированное тестирование	10
1.3 Сценарии использования	11
1.4 Определение требований к системе	11
2 Проектирование системы, разработка языка и алгоритма	12
2.1 Архитектура системы	12
ЗАКЛЮЧЕНИЕ	16
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	17
ПРИЛОЖЕНИЕ А Исходный код	18

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящей выпускной квалификационной работе бакалавра применяют следующие термины с соответствующими определениями:

Перенос данных — это процесс копирования и перемещения данных из одного источника или системы в другую без удаления данных из оригинальной базы данных

DevOps-инженер — ...

CI/CD — ...

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящей выпускной квалификационной работе бакалавра применяют следующие сокращения и обозначения:

БД — база данных

СУБД — система управления базами данных

ВВЕДЕНИЕ

В современном мире рынок онлайн-образования переживает бурное развитие. Всё больше людей всех возрастов обращаются к дистанционным форматам, которые предлагают гибкость и доступность обучения в любой точке земного шара. Сегмент онлайн-образования, ориентированного на школьников, занимает значительную долю рынка. [1] В связи с этим особого внимания заслуживают платформы, предназначенные для школьников, их родителей и учителей.

Такие системы позволяют интегрировать учебные материалы, задания, оценки и обратную связь в едином пространстве. Для школьников это возможность учиться в интерактивной и увлекательной форме, для родителей — возможность отслеживать успехи своих детей и принимать участие в образовательном процессе, а для учителей — эффективное управление учебным процессом и персонализированное обучение.

Эти платформы требуют надежной и эффективной системы управления данными для обеспечения бесперебойного функционирования и поддержки большого количества пользователей. Одним из популярных решений в области управления базами данных является PostgreSQL [2] — реляционная система управления базами данных с открытым исходным кодом. PostgreSQL завоевала доверие разработчиков благодаря своей надежности, богатому функционалу и способности обрабатывать большие объемы данных.

В процессе разработки платформы важно не только работать над техническими аспектами, но и уделять особое внимание качеству и надежности. Это делает тестирование незаменимой стадией разработки. Благодаря тестированию можно убедиться в стабильности и бесперебойной работе систем.

Использование реальных данных и проведение тестов в производственной среде позволяет максимально точно оценить работоспособность системы. Но такой подход сопряжён с рисками, такими как нарушения конфиденциальности данных, потенциальные сбои в работе системы и возможное негативное влияние на реальных пользователей.

Чтобы тестирование образовательных платформ было безопасным и в то же время эффективным, необходимо создать тестовую среду, максимально приближенную к производственным условиям. Создание тестовой среды

требует тщательной проработки и учета всех аспектов, связанных с функционированием платформы. Это включает в себя настройку инфраструктуры, идентичной производственной, и воспроизведение всех взаимосвязей данных, которые присутствуют в реальных условиях эксплуатации.

Особенно важна тут роль работы с данными. Для достижения максимальной схожести с производственной средой необходимо иметь тестовые данные, которые бы точно отражали объемы, структуры и взаимосвязи данных в реальной системе.

Создание тестового набора данных может осуществляться посредством различных методик. Один из подходов подразумевает полный перенос данных из производственной базы данных и дальнейшую анонимизацию конфиденциальной информации. Данную операцию можно выполнить с помощью утилиты `pg_dump` [3], которая обеспечивает экспорт данных. Для защиты пользовательской информации процесс анонимизации можно выполнять при помощи утилиты `pg_anonymizer` [4]. Несмотря на эффективность выбора подобных средств, основным недостатком данного метода выступает невысокая скорость переноса данных.

Иногда базы данных в крупных программных продуктах могут содержать терабайты данных [5]. В таких случаях процесс полного переноса данных способен занимать продолжительное время. Однако в ряде ситуаций столь длительное время, затрачиваемое на создание тестовых данных, является недопустимым.

Как правило, для покрытия большинства тестовых сценариев требуется не полный объём данных, а лишь их часть, что указывает на необходимость частичного переноса данных. Однако данный подход вызывает несколько вопросов.

В случае полного переноса тестирующему предоставляется доступ ко всему массиву данных, на основе которого он может проводить тестовые сценарии. При частичном переносе данных тестирующему необходимо каким-то образом определить и описать те данные, которые ему требуются, желательно не углубляясь в детали структуры базы данных и существующие в ней связи. Таким образом, первый вопрос заключается в следующем: каким образом тестирующему эффективно описать данные, необходимые для его работы?

Следующим важным вопросом является обеспечение целостности данных. При полном переносе реляционная целостность соблюдается автоматически, так как все взаимосвязанные данные переносятся в полном объёме, обеспечивая корректность и согласованность системы. Однако при частичном переносе необходимо серьезное внимание уделить поддержанию целостности данных. Недопустимо, чтобы после переноса в тестовую среду данных оказалось недостаточно для выполнения тестовых сценариев. Следовательно, второй вопрос заключается в следующем: как сохранить целостность данных, минимизируя при этом объём переносимых данных?

Перенос данных из производственной среды с использованием анонимизации является лишь одним из методов получения тестовых данных. Альтернативный метод заключается в генерации данных на основе заданных характеристик. Как может выглядеть такой сценарий? Пользователь определяет характеристики, такие как возраст, пол или средний балл по математике, при этом каждый параметр может сопровождаться набором или диапазоном значений. Далее программа генерирует тестовые данные, опираясь на сами данные, либо схему данных из производственной базы, а также на заданные пользователем характеристики. В результате полученные данные будут иметь структуру, аналогичную реальным данным, однако они не будут ассоциироваться с какими-либо конкретными записями из производственной базы. Тем не менее, для данного подхода также встают два ключевых вопроса: как пользователю описать требуемые данные и как их корректно сгенерировать.

Дипломная работа будет посвящена разработке системы, которая на основе пользовательских запросов обеспечит перенос взаимосвязанных данных из одной базы данных в другую с применением методов анонимизации, а также будет способна генерировать тестовые данные. В рамках исследования будут рассмотрены алгоритмы переноса и генерации данных, а также разработка языка для описания данных, необходимых пользователю.

1 Постановка задачи и теоретические предпосылки

1.1 Тестирование

Чтобы определить требования к нашей системе, для начала нужно разобраться с основными аспектами тестирования программного обеспечения.

1.1.1 Процесс тестирования

Тестирование — это процесс оценки системы или её компонентов с целью проверки их соответствия заданным требованиям. Основная цель тестирования заключается в выявлении дефектов, обеспечении качества и подтверждении работоспособности программного обеспечения.

Процесс тестирования можно разбить на несколько ключевых этапов [6]:

а) Планирование:

- разрабатывается стратегия тестирования, включая выбор типов тестирования;
- создается тест-план, где описываются ресурсы, сроки, и критерии начала и завершения тестирования;
- проводится анализ рисков, связанных с тестированием и определяются действия для минимизации их воздействия.

б) Подготовка:

- разработка тестовых случаев и сценариев на основе требований и спецификаций;
- подготовка тестовой среды, включая оборудование и программные средства, необходимые для проведения тестов;
- подготовка данных для тестирования и настройка автоматизированных тестов, если они предусмотрены.

в) Проведение:

- непосредственное выполнение тестов согласно плану тестирования;
- фиксация и документирование результатов тестирования, выявление дефектов и несоответствий.

г) Совершенствование:

- анализ результатов тестирования и составление отчёта по итогам выполненной работы;
- обмен опытом и полученными знаниями внутри команды и с другими заинтересованными сторонами;
- обновление и улучшение тестовых документов и процессов на основе полученных данных.

Наша система будет применяться на этапе подготовки: с её помощью можно будет создавать тестовые данные.

1.2 Ручное и автоматизированное тестирование

Тестирование можно классифицировать на ручное и автоматизированное.

Ручное тестирование предполагает процесс, при котором тестировщик выполняет тесты без помощи автоматизированных инструментов. Оно позволяет глубже вникнуть в пользовательский опыт, и его легко применять к новым или часто изменяющимся функциональностям, однако оно может быть времяемким, подвержено человеческим ошибкам и не всегда позволяет повторить результаты для их сравнения.

Автоматизированное тестирование предполагает использование программных инструментов для автоматизации выполнения тестовых сценариев. Это предполагает разработку скриптов, которые автоматически выполняют тесты и сверяют результаты с ожидаемыми. Автоматизированное тестирование обеспечивает высокую скорость и стабильность результатов, что делает его эффективным для регрессионного тестирования и стабильных функциональностей, однако требует значительных ресурсов для разработки скриптов и не всегда подходит для тестирования пользовательских интерфейсов и динамически изменяющихся требований.

Оба подхода часто используются совместно для достижения наилучшего результата. Обычно ручное тестирование применяют на начальных стадиях или для исследовательского тестирования, тогда как автоматизированное тестирование используют для повторяющихся задач или сложных сценариев, которые требуют стабильного выполнения.

1.3 Сценарии использования

1.4 Определение требований к системе

2 Проектирование системы, разработка языка и алгоритма

2.1 Архитектура системы

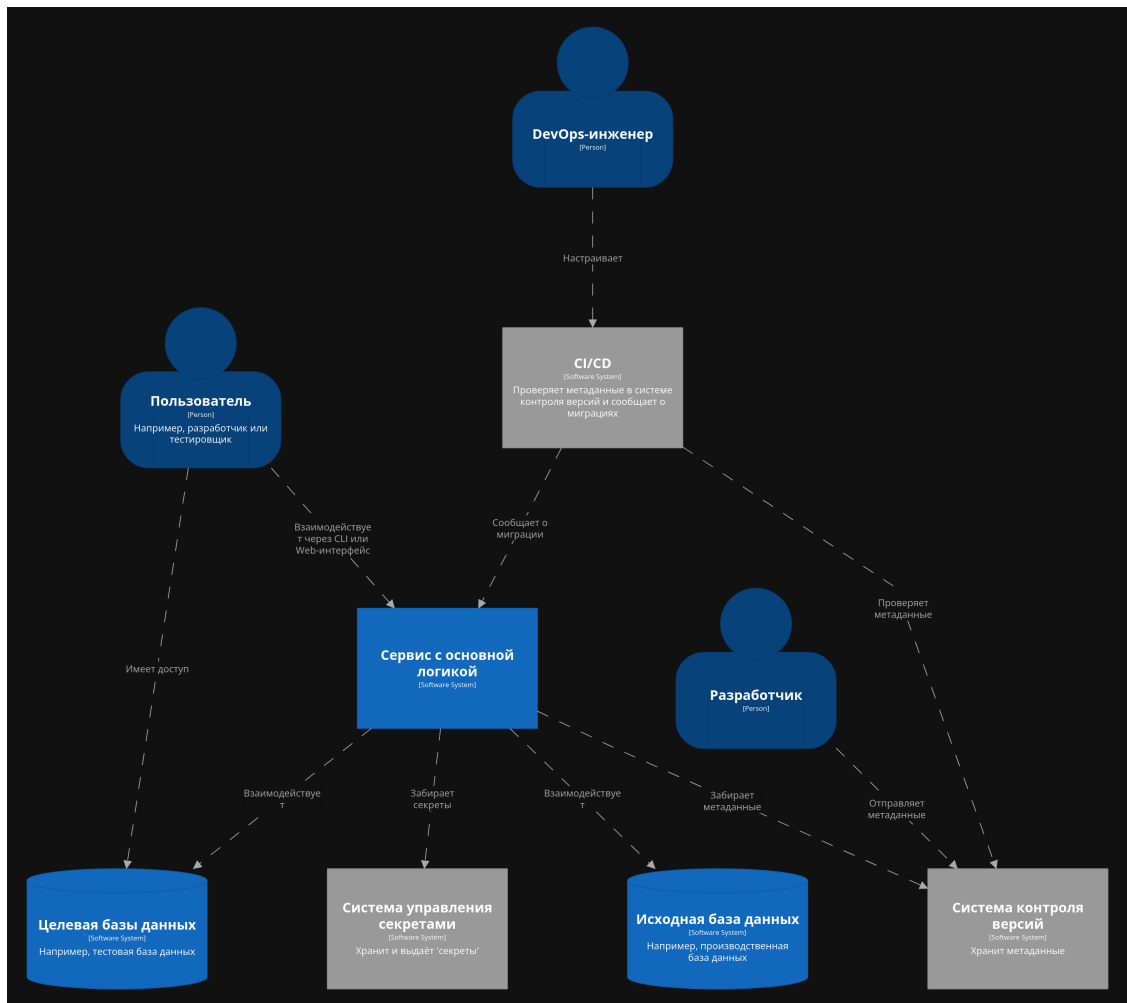


Рисунок 1 – System Context

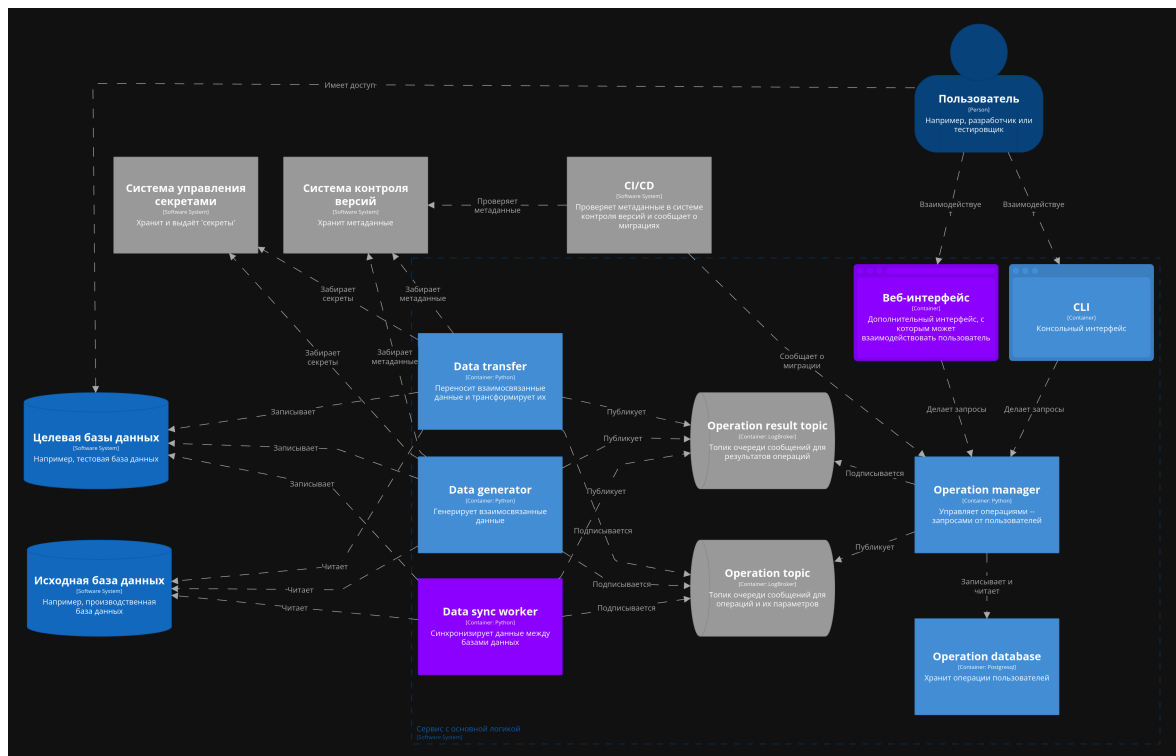


Рисунок 2 – Containers

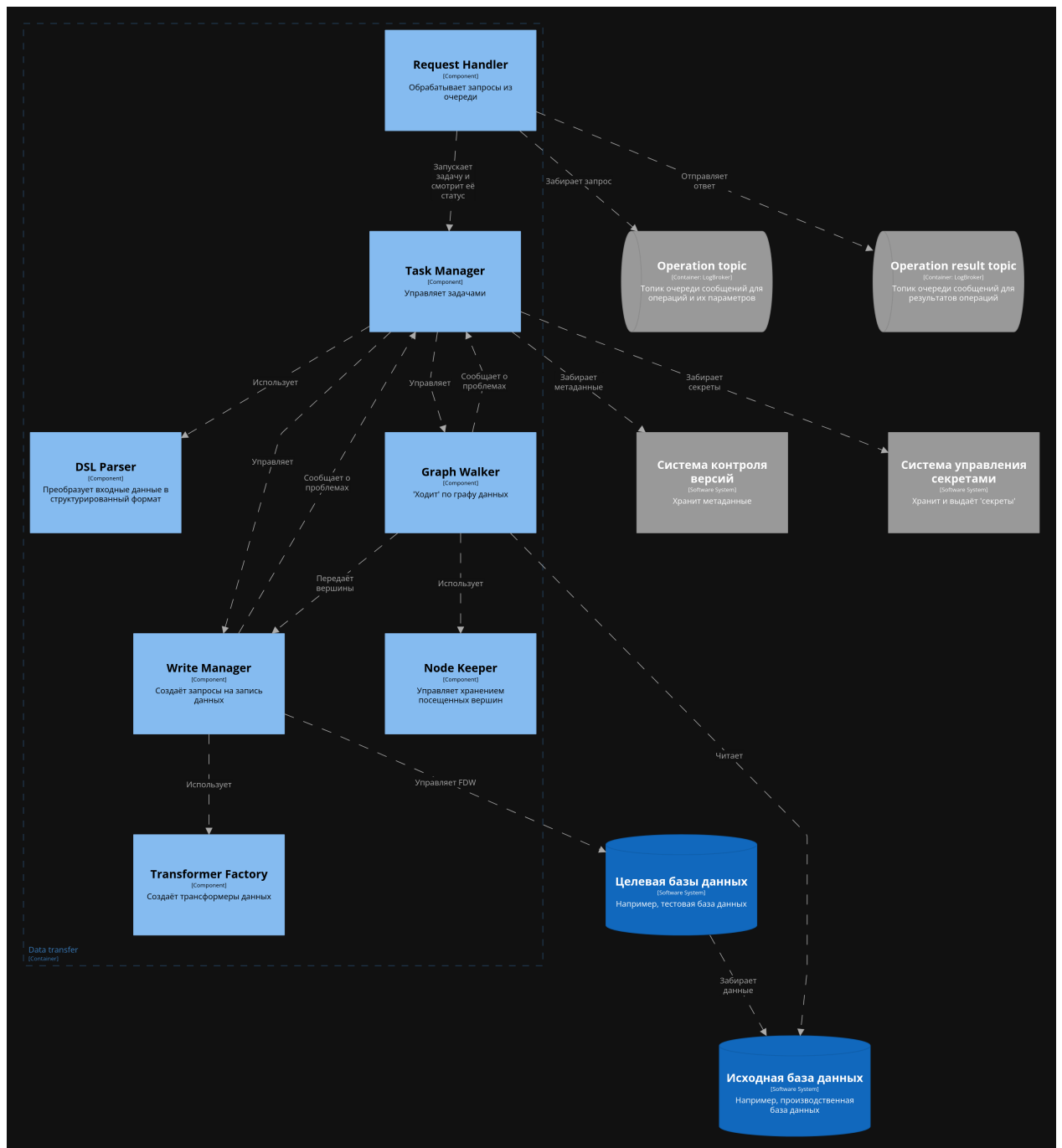


Рисунок 3 – Data Transfer Components

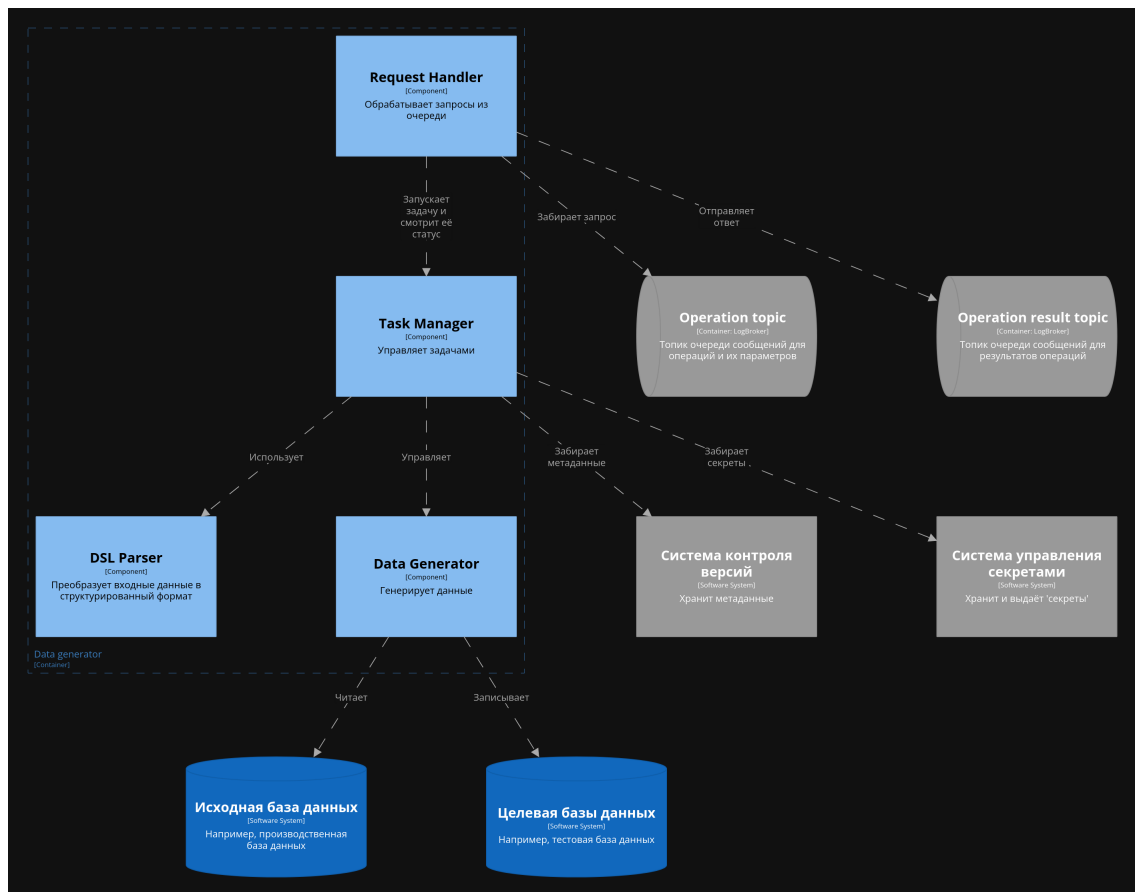


Рисунок 4 – Data Generator Components

ЗАКЛЮЧЕНИЕ

Здесь могло бы быть ваше заключение!

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Исследование рынка онлайн-образования. — URL: <https://main.talenttech.ru/research/issledovanie-rynka-onlajn-obrazovaniya/> (дата обращения 01.03.2025).
2. PostgreSQL: The world's most advanced open source database. — URL: <https://www.postgresql.org/> (дата обращения 02.01.2025).
3. PostgreSQL: Documentation: 17: pg_dump. — URL: <https://www.postgresql.org/docs/current/app-pgdump.html> (дата обращения 02.01.2025).
4. PostgreSQL: pg_anonymize, a new extension for simple and transparent data anonymization. — URL: https://www.postgresql.org/about/news/pg_anonymize-a-new-extension-for-simple-and-transparent-data-anonymization-2606/ (дата обращения 02.01.2025).
5. Очень большие таблицы в PostgreSQL. Или как превратить 60+ Tb в 10+ Tb | PgConf.Russia 2017 | PGConf.Russia. — URL: <https://pgconf.ru/talk/1588070> (дата обращения 02.01.2025).
6. Блэк Р. Ключевые процессы тестирования. — Лори, 2025. — ISBN 978-5-85582-392-9.

ПРИЛОЖЕНИЕ А

Исходный код

Здесь может быть будет qr-код или ссылка на репозиторий.