



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 12 страниц, 4 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
ЗАКЛЮЧЕНИЕ . . . . .	10
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .	11
ПРИЛОЖЕНИЕ А Исходный код . . . . .	12

## **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

## **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 образовательные платформы стали неотъемлемой частью учебного процесса, оказывая значительное влияние на эффективность образования и доступность знаний. С развитием цифровых технологий и интернета, образовательные ресурсы стали более доступными, что способствует росту интереса к онлайн-обучению среди самой широкой аудитории. Актуальность образовательных платформ заключается в их способности предоставить пользователям разнообразные инструменты и материалы для обучения независимо от географического положения и временных ограничений.

Особое внимание привлекают 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 [1] — реляционная система управления базами данных с открытым исходным кодом. PostgreSQL завоевала доверие разработчиков благодаря своей надежности, богатому функционалу и способности обрабатывать большие объемы данных.

PostgreSQL часто используется в образовательных платформах ввиду своей гибкости и масштабируемости. Эта СУБД обеспечивает необходимые инструменты для хранения и управления сложными и взаимосвязанными данными, которые активно используются в процессе обучения. Например, образовательные платформы могут использовать различные типы данных — от текстовых материалов до данных об успеваемости обучающихся. PostgreSQL позволяет эффективно управлять этими данными, обеспечивая высокую скорость доступа и надежность хранения.

Кроме того, использование PostgreSQL в образовательных платформах

поддерживает ожидания пользователей по быстродействию и доступности, что является критически важным в условиях интенсивного использования и необходимости в реальном времени предоставлять актуальные данные. Платформа на основе PostgreSQL может масштабироваться по мере роста числа пользователей и объема данных, что делает ее идеальным выбором в качестве СУБД.

Однако, помимо технического аспекта разработки, одной из ключевых задач является обеспечение качества и надежности платформы, что делает тестирование неотъемлемой частью процесса разработки. Тестирование образовательных платформ играет фундаментальную роль по нескольким причинам. Прежде всего, это гарантирует, что системы работают стабильно и без сбоев, предоставляя пользователям доступ к ресурсам и инструментам без задержек и ошибок. Безотказная работа особенно важна в учебном процессе, где сбои могут серьезно нарушить образовательную деятельность.

К тому же, тестирование способствует улучшению пользовательского опыта, делая взаимодействие с системой интуитивным и удовлетворяющим требования пользователей. Это включает удобство навигации, скорость отклика системы и общую функциональность интерфейса. Все это способствует более эффективному обучению и вовлеченности учащихся.

Таким образом, тестирование является жизненно важным этапом, обеспечивающим стабильно высокое качество образовательных платформ и повышающим их ценность для всех участников образовательного процесса.

Чтобы тестирование образовательных платформ было эффективным и всесторонним, необходимо создать тестовую среду, максимально приближенную к производственным условиям. Это не только позволяет выявить потенциальные ошибки и сбои на ранних стадиях разработки, но и помогает оценить реальную производительность и надежность системы в условиях, близких к тем, с которыми столкнутся конечные пользователи.

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. Такой подход помогает разработчикам и тестировщикам более точно предсказать, как платформа будет вести себя под различными нагрузками и в разных сценариях использования.

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. Тестовая база данных представляет собой набор, полученный из реальных данных, предназначенный для проверки заданных тестовых сценариев.

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` [2], которая обеспечивает экспорт данных. Для защиты пользовательской информации процесс анонимизации можно выполнять при помощи утилиты `pg_anonymizer` [3]. 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 [4]. 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 тестировщику предоставляется доступ ко всему массиву данных, на основе которого он может проводить тестовые сценарии. При частичном переносе данных тестировщику 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 тестировщику 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.

## **ЗАКЛЮЧЕНИЕ**

Здесь могло бы быть ваше заключение!

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. <https://www.postgresql.org/>. — URL: <https://www.postgresql.org/> (дата обращения 02.01.2025).
2. PostgreSQL: Documentation: 17: pg\_dump. — URL: <https://www.postgresql.org/docs/current/app-pgdump.html> (дата обращения 02.01.2025).
3. 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/](https://www.postgresql.org/about/news/pg_anonymize-a-new-extension-for-simple-and-transparent-data-anonymization-2606/) (дата обращения 02.01.2025).
4. <https://pgconf.ru/talk/1588070>. — URL: <https://pgconf.ru/talk/1588070> (дата обращения 02.01.2025).

## **ПРИЛОЖЕНИЕ А**

### **Исходный код**

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