МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ)

Рабочая программа дисциплины (модуля),
как компонент образовательной программы
базового высшего образования
по направлению подготовки
09.03.01 Информатика и вычислительная техника,
утвержденной первым проректором ФГАОУ ВО
РУТ (МИИТ) Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Базы данных и основы SQL

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): Цифровая инженерия транспортных процессов

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная ФГАОУ ВО РУТ (МИИТ)
ID подписи: 937226
Подписал: руководитель образовательной программы Проневич Ольга Борисовна
Дата: 02.09.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся системных знаний и практических навыков проектирования, администрирования и эффективного использования реляционных и нереляционных баз данных для решения задач хранения, предобработки и подачи данных в конвейеры машинного обучения.

Задачами освоения дисциплины являются:

- Формирование теоретических знаний об архитектурах систем управления базами данных, моделях данных и принципах нормализации для обеспечения целостности и производительности.

- Освоение практических навыков написания запросов на языке SQL для извлечения, агрегации и трансформации данных.

- Приобретение умений проектировать и разворачивать базы данных под задачи аналитики и хранения сырых данных.

- Формирование компетенций по оптимизации производительности запросов и мониторингу состояния СУБД.

- Освоение инструментов версионирования схем данных и интеграции баз данных с языком Python для создания процессов обработки данных.

- Приобретение навыков работы с нереляционными хранилищами для обработки слабоструктурированных данных.

Дисциплина предназначена для формирования компетенций, необходимых для решения профессиональных задач, связанных с обеспечением качественных данных для обучения моделей, управлением версиями данных и оптимизацией инфраструктуры хранения.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ППК-Р4 - Способен использовать базы данных при создании программных модулей и компонентов.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

- Архитектуры современных систем управления базами данных, включая реляционные и нереляционные модели данных, их сильные и слабые стороны для различных задач хранения и анализа данных.

- Синтаксис языка SQL, особенности программирования на нём, а также современные подходы к проектированию баз данных и инструменты для работы с ними.

- Внутреннее устройство СУБД, методы и средства мониторинга и оптимизации производительности для обеспечения эффективной работы с большими объёмами данных.

Уметь:

- Применять языки работы с базами данных для создания сложных аналитических выборок, включая оконные функции, обобщённые табличные выражения и подзапросы.

- Использовать современные среды программирования и библиотеки для гибкой работы с данными в базе, написания скриптов обработки и интеграции с инструментами анализа данных.

- Применять методы и средства мониторинга производительности, вырабатывать варианты оптимизации запросов и обосновывать выбранные решения.

Владеть:

- Навыками написания запросов на языке SQL для извлечения, агрегации, трансформации и очистки данных, а также создания и управления объектами базы данных.

- Навыками работы с нереляционными хранилищами для сбора и обработки слабоструктурированных данных.

- Навыками интеграции баз данных с языком программирования для создания пайплайнов обработки данных и взаимодействия с моделями машинного обучения.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов		
	Всего	Семестр	
		№1	№2

Контактная работа при проведении учебных занятий (всего):	64	32	32
В том числе:			
Занятия лекционного типа	32	16	16
Занятия семинарского типа	32	16	16

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 80 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	Тема 1. Введение в архитектуру баз данных для Data Science. Рассматриваемые вопросы: - Роль систем управления базами данных в жизненном цикле машинного обучения. - Классификация СУБД: реляционные, нереляционные, NewSQL. - Обзор рынка СУБД: PostgreSQL, MySQL, ClickHouse, MongoDB. - Базовые понятия: транзакция, ACID, CAP-теорема. - Выбор СУБД в зависимости от задач хранения и анализа данных. - Инструментарий для работы с БД: обзор клиентских приложений.
2	Тема 2. Реляционная модель данных. Рассматриваемые вопросы: - Основы реляционной модели: отношения, кортежи, атрибуты. - Первичный и внешний ключи. - Целостность данных: сущностная и ссылочная. - Типы данных в реляционных базах. - Сравнение реляционной модели с другими подходами.
3	Тема 3. Язык SQL: DDL и DML. Рассматриваемые вопросы: - Стандарт SQL и его диалекты. - Типы данных: числовые, строковые, дата/время, JSON/JSONB. - Создание, изменение и удаление таблиц (CREATE, ALTER, DROP). - Ограничения целостности: первичный и внешний ключи, уникальность, проверка. - Вставка, обновление и удаление данных (INSERT, UPDATE, DELETE).

№ п/п	Тематика лекционных занятий / краткое содержание
4	Тема 4. Базовые операции выборки данных. Рассматриваемые вопросы: - Синтаксис SELECT: выборка столбцов, фильтрация (WHERE), сортировка (ORDER BY). - Логические операторы и операторы сравнения. - Работа с NULL-значениями. - Ограничение числа записей (LIMIT, OFFSET). - Практические примеры выборок для подготовки данных.
5	Тема 5. Сложные выборки данных. Рассматриваемые вопросы: - Фильтрация и сортировка данных. - Группировка данных и агрегатные функции (GROUP BY, HAVING). - Внутренние и внешние соединения таблиц (JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN). - Подзапросы и их применение в аналитических задачах. - Практические примеры выборок для подготовки данных для ML.
6	Тема 6. Оконные функции и обобщённые табличные выражения. Рассматриваемые вопросы: - Конструкция OVER() и определение окна данных. - Ранжирующие функции: ROW_NUMBER, RANK, DENSE_RANK. - Функции смещения: LAG, LEAD. - Сравнение оконных функций с группировкой. - Обобщённые табличные выражения (CTE): рекурсивные и нерекурсивные. - Применение для задач аналитики и подготовки данных для ML.
7	Тема 7. Транзакции и уровни изоляции. Рассматриваемые вопросы: - Понятие транзакции и ее свойства ACID. - Управление транзакциями: BEGIN, COMMIT, ROLLBACK. - Уровни изоляции транзакций. - Фантомные чтения, неповторяемое чтение, грязное чтение. - Применение транзакций в ETL-процессах.
8	Тема 8. Проектирование реляционных баз данных. Рассматриваемые вопросы: - Сущности, атрибуты, ключи. - Связи между таблицами: один-к-одному, один-ко-многим, многие-ко-многим. - Нормализация: первая, вторая, третья нормальные формы. - Нормальная форма Бойса-Кодда. - Денормализация для задач аналитики и хранилищ данных. - Построение ER-диаграмм.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Установка и настройка среды для работы с базами данных. Рассматриваемые вопросы: - Установка и настройка СУБД (PostgreSQL). - Установка инструментов администрирования (pgAdmin). - Создание персонального имени входа и рабочей базы данных.

№ п/п	Тематика практических занятий/краткое содержание
	- Настройка окружения для работы с базами данных. - Проверка работоспособности и подключение к серверу.
2	Тема 2. Создание и управление структурами данных. Рассматриваемые вопросы: - Создание таблиц с учётом типов данных и ограничений. - Модификация структуры таблиц (ALTER TABLE). - Создание внешних ключей для обеспечения ссылочной целостности. - Построение диаграммы базы данных в pgAdmin. - Проектирование структуры базы данных по заданной предметной области.
3	Тема 3. Наполнение и изменение данных. Рассматриваемые вопросы: - Вставка данных с учётом зависимостей между таблицами (INSERT). - Обновление и удаление записей (UPDATE, DELETE). - Тестирование каскадных действий при удалении и обновлении. - Реализация базовых операций в рамках заданного сценария. - Работа с редактором данных в pgAdmin.
4	Тема 4. Базовые операции выборки данных. Рассматриваемые вопросы: - Формирование запросов с фильтрацией (WHERE). - Сортировка данных (ORDER BY). - Ограничение числа записей (LIMIT, OFFSET). - Работа с NULL-значениями. - Практика на транспортных данных.
5	Тема 5. Сложные выборки данных: соединения и подзапросы. Рассматриваемые вопросы: - Формирование запросов с фильтрацией, сортировкой и ограничением. - Выборка данных из нескольких таблиц с использованием различных видов соединений (JOIN). - Применение подзапросов для решения практических задач. - Реализация пакета команд. - Практика на транспортных данных.
6	Тема 6. Группировка и агрегация данных. Рассматриваемые вопросы: - Получение агрегированных показателей с использованием группировки (GROUP BY). - Фильтрация с использованием агрегированных значений (HAVING). - Применение группировки для решения аналитических задач. - Сравнение различных подходов к агрегации данных. - Практика на транспортных данных.
7	Тема 7. Оконные функции на практике. Рассматриваемые вопросы: - Решение задач с использованием оконных функций (ROW_NUMBER, RANK, DENSE_RANK). - Использование функций смещения (LAG, LEAD). - Сравнение оконных функций с традиционными подходами. - Применение оконных функций для анализа транспортных данных.
8	Тема 8. Обобщённые табличные выражения (CTE). Рассматриваемые вопросы: - Применение обобщённых табличных выражений для упрощения сложных запросов. - Рекурсивные CTE для работы с иерархическими данными. - Не рекурсивные CTE для структурирования запросов. - Практика на транспортных данных.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с лекционным материалом и дополнительной литературой.
2	Изучение документации по системам управления базами данных
3	Подготовка к практическим занятиям и выполнение домашних заданий.
4	Решение задач по проектированию баз данных и написанию сложных запросов
5	Освоение системы контроля версий Git для управления версиями SQL-скриптов и схем данных.
6	Подготовка к промежуточной аттестации и текущему контролю.
7	Подготовка к промежуточной аттестации.
8	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Базы данных и основы языка SQL : учебное пособие / составитель С. В. Коломийцева. — Хабаровск : ДВГУПС, 2023. — 89 с.	https://e.lanbook.com/book/433529
2	Гудов, А. М. Базы данных и системы управления базами данных. Программирование на языке PL/SQL : учебное пособие / А. М. Гудов, С. Ю. Завозкин, Т. С. Рейн. — Кемерово : КемГУ, 2010. — 133 с. — ISBN 978-5-8353-1005-0.	https://e.lanbook.com/book/30135

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

MS Office (Excel, Word)

Notepad++

Браузер Chrome

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

Python 3.11/3.12 и JupyterLab / Jupyter Notebook

pandas, NumPy, SciPy, statsmodels и Pingouin

Matplotlib, Seaborn и Plotly

PostgreSQL и DBeaver Community Edition
Microsoft Excel и Power BI Desktop / LibreOffice Calc
Yandex DataLens, Metabase или Apache Superset
Git и GitHub / GitLab

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Компьютер преподавателя с доступом к сети Интернет
Компьютеры студентов с установленным Python-окружением и табличным процессором
Доступ к учебной базе событийных данных и экспериментальным выборкам
Доступ к PostgreSQL и BI-системе для визуализации результатов
Доступ к облачной виртуальной машине или Docker-контейнерам для воспроизводимого анализа
Проектор и экран для проектора
Маркерная доска

9. Форма промежуточной аттестации:

Зачет в 1, 2 семестрах.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

руководитель образовательной программы О.Б. Проневич

Согласовано:

Директор Д.В. Паринов

Руководитель образовательной программы О.Б. Проневич

Председатель учебно-методической комиссии Д.В. Паринов