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)

Архитектура data-продуктов для транспорта

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: 01.09.2026

1. Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются формирование у обучающихся системных знаний и практических навыков в области проектирования архитектуры data-продуктов для транспортной отрасли. В рамках курса рассматриваются современные подходы к построению архитектуры данных, включая выбор технологического стека, проектирование систем хранения и обработки данных, организацию потоков данных, внедрение систем мониторинга качества данных и построение информационных панелей. Особое внимание уделяется практическим аспектам расследования первопричин проблем с данными, выбору инструментов мониторинга и учету возможностей существующей программно-технической архитектуры при проектировании data-продуктов.

Задачами освоения дисциплины являются:

- формирование теоретических знаний о принципах построения архитектуры data-продуктов, включая выбор технологического стека и интеграцию с существующей программно-технической архитектурой;
- освоение инструментов мониторинга качества данных и принципов построения информационных панелей для визуализации показателей качества;
- развитие практических навыков расследования первопричин проблем с данными и разработки мероприятий по их устранению;
- формирование компетенций по проектированию и внедрению систем мониторинга качества данных для транспортных data-продуктов;
- освоение подходов к разработке требований к программному обеспечению и проектированию data-продуктов для транспортной отрасли.

Дисциплина предназначена для формирования компетенций, необходимых для эффективного проектирования, разработки и внедрения data-продуктов в транспортной отрасли с учетом требований к качеству данных и мониторингу.

2. Планируемые результаты обучения по дисциплине (модулю).

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

МП-ДТМ-1 - Способен управлять технологическими решениями в продуктовом контексте;

ОПК-6 - Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения;

УНК-05 - Способен проводить научно-исследовательские и опытно-конструкторские работы.

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

Знать:

- Инструменты мониторинга качества данных (Prometheus, Grafana, Great Expectations, Deequ) и принципы построения информационных панелей для визуализации показателей качества данных.
- Возможности существующей программно-технической архитектуры для реализации data-продуктов, включая системы хранения, обработки и оркестрации данных.
- Архитектурные подходы к построению data-продуктов (Data Lake, Data Warehouse, Lakehouse, Data Mesh), принципы проектирования систем обработки и хранения данных, методы разработки требований к программному обеспечению.

Уметь:

- Проектировать и внедрять системы мониторинга качества данных для транспортных data-продуктов, включая настройку инструментов и построение информационных панелей.
- Расследовать первопричины проблем с данными с использованием методологии root cause analysis, документировать результаты и разрабатывать корректирующие мероприятия.
- Разрабатывать требования к data-продуктам с учетом возможностей существующей программно-технической архитектуры, выбирать технологический стек и проектировать системы хранения и обработки данных.

Владеть:

- Навыками работы с инструментами мониторинга качества данных и построения информационных панелей для визуализации показателей качества.
- Методами расследования проблем с данными и документирования результатов, включая разработку корректирующих и предупреждающих мероприятий.
- Подходами к проектированию архитектуры data-продуктов, разработке требований и проектированию программного обеспечения с учетом архитектурных ограничений.

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 №7
Контактная работа при проведении учебных занятий (всего):	96	96
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	64	64

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 48 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 data-продуктов. Рассматриваемые вопросы: - Понятие data-продукта и его жизненный цикл. - Специфика транспортных данных для data-продуктов. - Классификация data-продуктов (аналитические, операционные, прогнозные). - Обзор современных подходов к архитектуре данных.

№ п/п	Тематика лекционных занятий / краткое содержание
2	Тема 2. Архитектурные подходы: Data Lake, Data Warehouse, Lakehouse, Data Mesh. Рассматриваемые вопросы: - Data Lake, Data Warehouse, Lakehouse: концепции, преимущества и ограничения. - Data Mesh: принципы и реализация. Data Fabric. - Сравнительный анализ подходов. - Выбор архитектуры для транспортного data-продукта.
3	Тема 3. Технологический стек для data-продуктов на транспорте. Рассматриваемые вопросы: - Хранилища данных (PostgreSQL, ClickHouse, Greenplum). - Инструменты оркестрации (Apache Airflow). - Инструменты обработки потоков (Apache Kafka, Spark Streaming). - Выбор технологического стека с учетом существующей архитектуры.
4	Тема 4. Проектирование систем хранения данных. Рассматриваемые вопросы: - Выбор модели хранения для транспортных данных. - Партиционирование и шардирование. - Сжатие и оптимизация хранения. - Управление жизненным циклом данных.
5	Тема 5. Организация потоков данных в data-продуктах. Рассматриваемые вопросы: - Потоковая и пакетная обработка данных. - Паттерны интеграции данных (CDC, ETL, ELT). - Маршрутизация и трансформация потоков. - Обеспечение согласованности данных.
6	Тема 6. Инструменты мониторинга и проектирование систем контроля качества. Рассматриваемые вопросы: - Обзор инструментов мониторинга (Prometheus, Grafana, Great Expectations, Deequ). - Цели и задачи системы мониторинга качества данных. - Метрики качества данных для мониторинга. - Настройка алертов и уведомлений.
7	Тема 7. Построение информационных панелей для качества данных. Рассматриваемые вопросы: - Принципы построения панелей для качества данных. - Визуализация показателей качества данных. - Интерактивные панели для анализа качества. - Настройка дашбордов в Grafana / Superset.
8	Тема 8. Data Governance и управление качеством данных. Рассматриваемые вопросы: - Понятие Data Governance и его роль в data-продуктах. - Политики управления данными и стандарты качества. - Обеспечение согласованности и достоверности данных. - Роль Data Steward в транспортных data-продуктах.
9	Тема 9. Расследование первопричин проблем с данными. Рассматриваемые вопросы: - Методология root cause analysis для проблем с данными. - Инструменты и методы расследования. - Документирование результатов расследования. - Разработка корректирующих и предупреждающих мероприятий.
10	Тема 10. Разработка требований к data-продуктам. Рассматриваемые вопросы:

№ п/п	Тематика лекционных занятий / краткое содержание
	- Методы сбора и анализа требований. - Структура требований к data-продукту. - Учет возможностей существующей программно-технической архитектуры. - Согласование требований с заинтересованными сторонами.
11	Тема 11. Проектирование программного обеспечения для data-продуктов. Рассматриваемые вопросы: - Архитектурные паттерны для data-продуктов. - Проектирование API и интерфейсов. - Микросервисная архитектура vs монолит. - Обеспечение масштабируемости и отказоустойчивости.
12	Тема 12. Безопасность data-продуктов на транспорте. Рассматриваемые вопросы: - Требования к безопасности данных на транспорте. - Управление доступом и аутентификация. - Шифрование данных. - Обеспечение приватности данных (152-ФЗ).

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Введение в архитектуру data-продуктов. Обзор инструментов. Рассматриваемые вопросы: - Обзор инструментов для создания data-продуктов. - Установка и настройка среды. - Знакомство с датасетами транспортной отрасли.
2	Тема 2. Выбор архитектуры для транспортного data-продукта. Рассматриваемые вопросы: - Анализ бизнес-задачи. - Выбор архитектуры (Data Lake, DWH, Lakehouse). - Обоснование выбора.
3	Тема 3. Проектирование базы данных для транспортного data-продукта. Рассматриваемые вопросы: - Выбор СУБД. - Разработка схемы базы данных. - Создание таблиц и связей.
4	Тема 4. Проектирование базы данных для транспортного data-продукта. Рассматриваемые вопросы: - Выбор СУБД. - Разработка схемы базы данных. - Создание таблиц и связей.
5	Тема 5. Настройка хранилища данных для транспортных данных. Рассматриваемые вопросы: - Установка и настройка ClickHouse/PostgreSQL. - Настройка партиционирования и индексов. - Оптимизация хранения.
6	Тема 6. Анализ существующей программно-технической архитектуры. Рассматриваемые вопросы:

№ п/п	Тематика практических занятий/краткое содержание
	- Анализ текущей архитектуры транспортной компании. - Выявление ограничений и возможностей. - Формирование требований к новому data-продукту с учетом существующей архитектуры.
7	Тема 7. Выбор технологического стека для data-продукта. Рассматриваемые вопросы: - Сравнение СУБД и инструментов оркестрации. - Выбор стека для конкретной задачи. - Обоснование выбора.
8	Тема 8. Проектирование системы хранения и обработки данных. Рассматриваемые вопросы: - Разработка архитектурной схемы data-продукта. - Определение компонентов и их взаимодействия. - Документирование архитектурного решения.
9	Тема 9. Организация ETL-пайплайна для транспортных данных. Рассматриваемые вопросы: - Создание ETL-пайплайна на Python. - Извлечение данных из источников. - Трансформация и загрузка данных
10	Тема 10. Знакомство с Apache Airflow. Рассматриваемые вопросы: - Установка и настройка Airflow. - Создание простого DAG. - Запуск и мониторинг задач.
11	Тема 11. Разработка DAG для ETL-пайплайна. Рассматриваемые вопросы: - Создание DAG для загрузки транспортных данных. - Настройка зависимостей и расписания. - Логирование и обработка ошибок.
12	Тема 12. Работа с потоковыми данными (Apache Kafka). Рассматриваемые вопросы: - Установка и настройка Kafka. - Создание топиков. - Публикация и потребление сообщений.
13	Тема 13. Инструменты мониторинга: установка и настройка. Рассматриваемые вопросы: - Установка Prometheus и Grafana. - Настройка источников данных. - Первые шаги в Grafana.
14	Тема 14. Мониторинг качества данных с Great Expectations. Рассматриваемые вопросы: - Установка и настройка Great Expectations. - Создание ожиданий (Expectations). - Проверка данных на соответствие.
15	Тема 15. Мониторинг качества данных с Deequ. Рассматриваемые вопросы: - Установка и настройка Deequ. - Создание профилей качества. - Проверка данных на соответствие метрикам.

№ п/п	Тематика практических занятий/краткое содержание
16	Тема 16. Визуализация метрик качества данных в Grafana.Тема 16. Визуализация метрик качества данных в Grafana. Рассматриваемые вопросы: - Создание панели для отображения метрик качества. - Визуализация показателей полноты, точности, уникальности. - Настройка дашборда.
17	Тема 17. Создание информационной панели для качества данных. Рассматриваемые вопросы: - Проектирование дашборда для мониторинга качества. - Интеграция с Prometheus. - Настройка обновления данных.
18	Тема 18. Настройка алертов и уведомлений в системе мониторинга. Рассматриваемые вопросы: - Настройка алертов в Grafana. - Настройка уведомлений. - Тестирование системы оповещения.
19	Тема 19. Расследование первопричин проблем с данными (практикум). Рассматриваемые вопросы: - Анализ данных с аномалиями. - Применение методов root cause analysis. - Документирование результатов.
20	Тема 20. Разработка требований к data-продукту. Рассматриваемые вопросы: - Анализ бизнес-потребностей. - Разработка требований к data-продукту. - Согласование требований с заинтересованными сторонами.
21	Тема 21. Проектирование API для data-продукта. Рассматриваемые вопросы: - Определение функций API. - Проектирование эндпоинтов. - Документирование API.
22	Тема 22. Реализация API для data-продукта. Рассматриваемые вопросы: - Реализация API с использованием FastAPI/Flask. - Настройка авторизации. - Тестирование API.
23	Тема 23. CI/CD для data-продуктов (GitHub Actions). Рассматриваемые вопросы: - Настройка GitHub Actions для проекта. - Автоматизация сборки и тестирования. - Развертывание data-продукта.
24	Тема 24. Обеспечение безопасности data-продукта. Рассматриваемые вопросы: - Настройка аутентификации и авторизации. - Шифрование данных. - Обеспечение соответствия требованиям 152-ФЗ.
25	Тема 25. Масштабирование data-продукта. Рассматриваемые вопросы: - Анализ производительности.

№ п/п	Тематика практических занятий/краткое содержание
	- Оптимизация запросов. - Горизонтальное масштабирование.
26	Тема 26. Интеграция data-продукта с существующей архитектурой. Рассматриваемые вопросы: - Анализ существующей архитектуры. - Разработка плана интеграции. - Реализация интеграции.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение литературы по архитектуре data-продуктов и системам мониторинга качества данных.
2	Работа с документацией инструментов (Apache Airflow, Prometheus, Grafana, Great Expectations, Deequ).
3	Выполнение домашних заданий по проектированию архитектуры, настройке мониторинга и построению панелей.
4	Самостоятельное изучение кейсов data-продуктов в транспортной отрасли (подготовка обзора 2-3 кейсов).
5	Анализ существующих архитектурных решений и подготовка рекомендаций по их улучшению.
6	Выполнение курсовой работы.
7	Подготовка к промежуточной аттестации.
8	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

- Проектирование архитектуры Data Lake для агрегации данных о движении подвижного состава железнодорожного транспорта.

- Разработка системы мониторинга качества данных для прогнозной модели пассажиропотоков городского транспорта.

- Выбор и обоснование технологического стека (ClickHouse, Airflow, Kafka) для операционной аналитики логистической компании.

- Проектирование хранилища данных (Data Warehouse) для учета и анализа использования топлива парком транспортных средств.

- Внедрение Data Mesh-подхода в транспортной компании: декомпозиция по доменам и проектирование интерфейсов обмена данными.

- Построение информационной панели (Grafana/Superset) для визуализации метрик качества данных телематических систем.

- Расследование первопричин (Root Cause Analysis) инцидентов с данными в системе управления перевозками и разработка корректирующих мероприятий.

- Проектирование ETL/ELT-пайплайна для интеграции данных из разнородных источников (GPS-трекеры, АСУ, билетные системы) в единое хранилище.

- Разработка требований к data-продукту для прогнозирования загрузки вагонов с учетом ограничений существующей программно-технической архитектуры.

- Обеспечение безопасности и соответствия 152-ФЗ при проектировании архитектуры data-продукта для обработки персональных данных пассажиров.

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

№ п/п	Библиографическое описание	Место доступа
1	Силич, М. П. Моделирование и анализ бизнес-процессов : учебное пособие / М. П. Силич, В. А. Силич. — Москва : ТУСУР, 2011. — 213 с. — ISBN 978-5-86889-511-1.	https://e.lanbook.com/book/11794
2	Моделирование и анализ бизнес-процессов : учебное пособие / составители Т. В. Галанина, М. И. Баумгартэн. — Кемерово : КузГТУ имени Т.Ф. Горбачева, 2023. — 164 с. — ISBN 978-5-00137-431-2.	https://e.lanbook.com/book/399725

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 7 семестре.
Курсовая работа в 7 семестре.

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