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)

Управление жизненным циклом данных

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: 09.09.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся системных знаний и практических навыков управления полным жизненным циклом данных в корпоративной среде, включая политики сбора, хранения, обработки, обеспечения качества, архивации, удаления и соблюдения нормативных требований, что обеспечивает целостность, доступность и безопасность данных транспортного предприятия на всех этапах их существования.

Дисциплина предназначена для формирования и закрепления у обучающихся компетенций, необходимых для решения следующих профессиональных задач:

Проектирование и внедрение политик управления жизненным циклом данных.

Реализация отказоустойчивых потоков обработки данных (batch, streaming).

Настройка мониторинга качества данных в ETL/ELT-процессах.

Управление версиями данных и обеспечение воспроизводимости.

Обеспечение соответствия нормативным требованиям (152-ФЗ, GDPR).

Оптимизация затрат на хранение данных на всех этапах жизненного цикла.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ИД-4 - Способен проектировать, развертывать и оптимизировать архитектуры хранилищ данных (DWH) и Data Lake.

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

Знать:

Паттерны обработки данных: пакетную (batch), потоковую (streaming), а также гибридные архитектуры Lambda и Kappa, их преимущества, ограничения и области применения в транспортной отрасли.

Методологию управления жизненным циклом данных: этапы жизненного цикла (создание, хранение, использование, архивация, удаление), классификацию данных по типам и критичности, принципы Data Governance

и Data Stewardship, нормативно-правовые требования к обращению с данными.

Уметь:

Разрабатывать отказоустойчивые потоки данных: проектировать ETL/ELT-пайплайны с учетом обработки ошибок, повторных попыток, гарантий доставки (exactly-once, at-least-once), использовать механизмы чекпоинтинга и восстановления после сбоев.

Настраивать мониторинг качества данных в ETL/ELT-процессах и потоках: определять метрики качества (полнота, точность, согласованность, актуальность), внедрять проверки на всех этапах обработки, настраивать алертинг при отклонениях, визуализировать показатели качества данных.

Владеть:

Навыками управления жизненным циклом данных в корпоративной среде, включая разработку политик хранения и архивации, управление версиями данных (Data Versioning), реализацию механизмов очистки и удаления данных в соответствии с регуляторными требованиями.

Навыками проектирования и реализации отказоустойчивых потоков обработки данных с использованием современных инструментов (Apache Kafka, Apache Spark, Apache Airflow), включая настройку мониторинга и обеспечение качества данных на всех этапах.

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 №6
Контактная работа при проведении учебных занятий (всего):	80	80
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	48	48

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 64 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 управление жизненным циклом данных. Рассматриваемые вопросы: - Понятие жизненного цикла данных (Data Lifecycle). - Этапы жизненного цикла: создание, хранение, использование, архивация, удаление. - Роль Data Governance в управлении данными. - Ключевые принципы управления данными в транспортной отрасли. - Обзор нормативных требований (152-ФЗ, GDPR).
2	Тема 2. Классификация и каталогизация данных. Рассматриваемые вопросы: - Типы данных: структурированные, полуструктурированные, неструктурированные. - Классификация по критичности и чувствительности. - Data Catalog и Data Discovery. - Построение карты данных (Data Mapping). - Инструменты для каталогизации данных.
3	Тема 3. Паттерны обработки данных: Batch-обработка. Рассматриваемые вопросы: - Архитектура пакетной обработки. - ETL/ELT-пайплайны. - Инструменты: Apache Airflow, Apache Spark (batch mode). - Преимущества и ограничения batch-подхода. - Кейсы использования в транспортной отрасли.
4	Тема 4. Паттерны обработки данных: Stream-обработка. Рассматриваемые вопросы: - Архитектура потоковой обработки. - Событийно-ориентированная архитектура. - Инструменты: Apache Kafka, Apache Flink, Apache Spark Streaming.

№ п/п	Тематика лекционных занятий / краткое содержание
	- Преимущества и ограничения streaming-подхода. - Применение для транспортных данных (GPS-треки, телеметрия).
5	Тема 5. Гибридные архитектуры: Lambda и Kappa. Рассматриваемые вопросы: - Архитектура Lambda: batch + speed слои. - Архитектура Kappa: unified streaming. - Сравнение и выбор архитектуры для транспортных задач. - Реализация на практике. - Примеры использования.
6	Тема 6. Управление качеством данных (Data Quality). Рассматриваемые вопросы: - Измерения качества данных: полнота, точность, согласованность, актуальность, уникальность. - Методы оценки качества. - Инструменты: Great Expectations, Deequ, Soda. - Интеграция проверок качества в ETL/ELT. - Мониторинг качества данных.
7	Тема 7. Мониторинг качества данных в потоках. Рассматриваемые вопросы: - Метрики качества для batch и streaming потоков. - Настройка алертинга при отклонениях. - Визуализация метрик качества. - Интеграция с системами мониторинга (Prometheus, Grafana). - Автоматическое реагирование на проблемы качества.
8	Тема 8. Управление версиями данных (Data Versioning). Рассматриваемые вопросы: - Необходимость версионирования данных. - Инструменты: DVC (Data Version Control), LakeFS. - Версионирование датасетов и моделей. - Воспроизводимость экспериментов. - Интеграция с Git и CI/CD.
9	Тема 9. Data Governance: политики и стандарты. Рассматриваемые вопросы: - Определение Data Governance. - Структура управления данными в организации. - Разработка политик и стандартов работы с данными. - Data Stewardship и ответственность за данные. - Внедрение Data Governance на транспорте.
10	Тема 10. Управление метаданными (Metadata Management). Рассматриваемые вопросы: - Типы метаданных: технические, бизнес, операционные. - Инструменты: Apache Atlas, Amundsen, OpenMetadata. - Сбор и хранение метаданных. - Использование метаданных для управления данными. - Data Lineage (происхождение данных).
11	Тема 11. Архивация и удаление данных. Рассматриваемые вопросы: - Стратегии архивации данных. - Tiered Storage: hot/warm/cold. - Автоматизация перемещения данных между уровнями.

№ п/п	Тематика лекционных занятий / краткое содержание
	- Политики удаления и очистки данных. - Обеспечение соответствия регуляторным требованиям.
12	Тема 12. Нормативно-правовое регулирование данных. Рассматриваемые вопросы: - 152-ФЗ "О персональных данных". - GDPR: требования к обработке данных. - Отраслевые нормативные акты на транспорте. - Обеспечение соответствия при проектировании систем. - Ответственность за нарушение требований.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Классификация данных транспортного предприятия. Рассматриваемые вопросы: - Анализ типов данных в транспортной компании. - Классификация по критичности и чувствительности. - Построение карты данных. - Разработка рекомендаций по категориям.
2	Тема 2. Разработка политик управления данными. Рассматриваемые вопросы: - Разработка политики хранения данных. - Политика качества данных. - Политика безопасности и доступа. - Политика архивации и удаления.
3	Тема 3. Построение ETL-пайплайна (Batch). Рассматриваемые вопросы: - Проектирование ETL-процесса для транспортных данных. - Реализация на Python с использованием Pandas. - Оркестрация с Apache Airflow. - Тестирование пайплайна.
4	Тема 4. Построение streaming-пайплайна с Kafka. Рассматриваемые вопросы: - Настройка Apache Kafka. - Создание producer для данных GPS-треков. - Создание consumer для обработки данных. - Реализация простого streaming-приложения.
5	Тема 5. Обработка streaming-данных с Spark Streaming. Рассматриваемые вопросы: - Настройка Spark Streaming. - Чтение данных из Kafka. - Агрегация и фильтрация потоковых данных. - Сохранение результатов.
6	Тема 6. Реализация Lambda-архитектуры. Рассматриваемые вопросы: - Проектирование Lambda-архитектуры для транспортного кейса. - Реализация batch и speed слоев.

№ п/п	Тематика практических занятий/краткое содержание
	- Объединение результатов. - Сравнение подходов.
7	Тема 7. Введение в Great Expectations. Рассматриваемые вопросы: - Установка и настройка Great Expectations. - Создание ожиданий (expectations) для данных. - Запуск проверок качества. - Генерация отчетов.
8	Тема 8. Интеграция проверок качества в ETL. Рассматриваемые вопросы: - Встраивание Great Expectations в ETL-пайплайн. - Проверка данных на всех этапах. - Обработка ошибок качества. - Логирование результатов.
9	Тема 9. Настройка мониторинга качества с Prometheus. Рассматриваемые вопросы: - Экспорт метрик качества в Prometheus. - Создание дашбордов в Grafana. - Настройка алертов. - Визуализация трендов качества.
10	Тема 10. Data Versioning с DVC. Рассматриваемые вопросы: - Установка и настройка DVC. - Версионирование датасетов. - Интеграция с Git. - Использование удаленных хранилищ.
11	Тема 11. Управление метаданными с OpenMetadata. Рассматриваемые вопросы: - Установка OpenMetadata. - Добавление источников данных. - Построение Data Lineage. - Использование каталога данных.
12	Тема 12. Реализация политик архивации данных. Рассматриваемые вопросы: - Проектирование системы архивации. - Автоматизация перемещения данных. - Настройка политик в облачном хранилище. - Тестирование архивации и восстановления.
13	Тема 13. Реализация Data Governance на практике. Рассматриваемые вопросы: - Разработка политик и процедур. - Назначение владельцев данных (Data Stewards). - Внедрение контроля доступа. - Документирование процессов.
14	Тема 14. Обеспечение соответствия 152-ФЗ. Рассматриваемые вопросы: - Анализ требований к системам с персональными данными. - Реализация мер защиты. - Документирование процессов обработки. - Проверка соответствия.

№ п/п	Тематика практических занятий/краткое содержание
15	Тема 15. Data Observability с Soda. Рассматриваемые вопросы: - Установка и настройка Soda. - Создание проверок на качество данных. - Интеграция с CI/CD. - Настройка уведомлений.
16	Тема 16. Оптимизация затрат на хранение. Рассматриваемые вопросы: - Анализ затрат на хранение. - Реализация Tiered Storage. - Настройка политик жизненного цикла в S3. - Оценка экономического эффекта.
17	Тема 17. Построение отказоустойчивого streaming-пайплайна. Рассматриваемые вопросы: - Обработка ошибок в streaming. - Реализация чекпоинтинга. - Гарантии доставки (exactly-once). - Тестирование восстановления после сбоя.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение официальной документации Apache Airflow, Apache Kafka, Great Expectations.
2	Самостоятельное проектирование ETL/ELT-пайплайна.
3	Разработка политик управления данными для транспортного предприятия.
4	Изучение материалов индустриальных партнеров (кейсы).
5	Подготовка к защите итогового проекта.
6	Подготовка к промежуточной аттестации.
7	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Панов, А. В. Управление жизненным циклом систем управления ресурсами и взаимоотношениями предприятия : учебно-методическое пособие / А. В. Панов. — Москва : РТУ МИРЭА, 2026. — 60 с. — ISBN 978-5-7339-2960-6.	https://e.lanbook.com/book/523317

2	Баланов, А. Н. E-commerce. Технологии и стратегии : учебное пособие для вузов / А. Н. Баланов. — Санкт-Петербург : Лань, 2024. — 472 с. — ISBN 978-5-507-49763-8.	https://e.lanbook.com/book/430130

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

https://habr.com/ru - база знаний в виде статей, обзоров

http://library.miit.ru/ - электронно-библиотечная система научно-технической библиотеки МИИТ

https://superset.apache.org/user-docs/intro/#documentation – документация Apache Superset

https://www.postgresql.org/ - официальная страница PostgreSQL

https://yandex.cloud/ru/docs/datalens/ - официальная документация Yandex Datalens

https://docs.docker.com/ - официальная документация Docker

https://www.metabase.com/docs/latest/ - официальная документация Metabase

https://tableau.pro/ - официальный сайт Tableau с лучшими практиками проектирования дашбордов

https://revealthedata.com/blog/ - Блог о визуализации данных и BI-системах

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

MS Office (Excel, Word, PowerPoint)
Apache Superset
Yandex Datalens
Metabase
PostgreSQL
Docker
Apache Airflow
Браузер Chrome / Яндекс браузер
Draw.io

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 6 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