

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
специализированного высшего образования
по направлению подготовки
25.04.03 Аэронавигация,
утвержденной первым проректором ФГАОУ ВО
РУТ (МИИТ) Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Сбор, хранение и обработка больших данных

Направление подготовки: 25.04.03 Аэронавигация

Направленность (профиль): Интеллектуальные системы обработки
информации и управления на воздушном
транспорте

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: 5665

Подписал: заведующий кафедрой Нутович Вероника
Евгеньевна

Дата: 01.09.2026

1. Общие сведения о дисциплине (модуле).

Дисциплина «Сбор, хранение и обработка больших данных» формирует у магистрантов системное видение архитектуры информационных потоков в авиатранспортной отрасли. В условиях тотального импортозамещения и перехода к интеллектуальным системам управления воздушным транспортом отрасль испытывает дефицит архитекторов данных, способных проектировать конвейеры обработки телеметрии, радарных данных и метеосводок. Студенты изучают глобальные стандарты обмена авиационной информацией, принципы построения озер данных и аналитических хранилищ, а также методы обеспечения качества данных для предиктивных моделей. Практическая часть курса построена вокруг сквозного кейса концептуального проектирования платформы больших данных, где обучающиеся выступают в роли системных архитекторов, формируя технические задания, контракты данных и архитектурные манифесты с опорой на отечественный ландшафт программного обеспечения.

Целью освоения дисциплины является формирование у обучающихся системных компетенций в области концептуального и логического проектирования архитектуры платформ больших данных для интеллектуальных систем управления на воздушном транспорте.

Для достижения поставленной цели в рамках дисциплины решается комплекс задач, направленных на формирование у обучающихся способности – анализировать гетерогенные источники авиационной телеметрии, проектировать многослойные хранилища данных, выбирать оптимальные технологии из реестра отечественного программного обеспечения, разрабатывать контракты данных и формировать архитектурную документацию для последующей реализации командами разработки.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ПК-1 - Способен управлять работами по созданию (модификации) и сопровождению интеллектуальных информационных систем, автоматизирующих задачи организационного управления и бизнес-процессы на воздушном транспорте.

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

Знать:

- парадигмы построения распределенных систем обработки больших данных и критерии их применимости в контурах управления воздушным транспортом;
- глобальные и отечественные стандарты обмена авиационной информацией, концепцию SWIM и форматы AIXM, FIXM, WXXM;
- принципы концептуального, логического и физического моделирования потоков данных с использованием нотаций UML, C4 и ER-моделей;
- архитектуру, роль и сценарии применения распределенных брокеров сообщений в IoT и телеметрических системах;
- форматы сериализации и колоночного хранения полуструктурированных данных и стратегии управления эволюцией схем;
- архитектурные паттерны организации озер данных, концепцию многоуровневой архитектуры «Медальон» и зоны приземления;
- стратегии партиципирования, шардирования и индексирования для оптимизации запросов к массивам пространственно-временных авиационных данных;
- архитектуру и отличительные особенности колоночных и транзакционных СУБД в ландшафте Big Data;
- парадигмы ETL и ELT, архитектурные различия и критерии выбора подхода в зависимости от вычислительных ресурсов и объема данных;
- концепции распределенных вычислений и in-memory обработки на уровне архитектурных возможностей;
- принципы оркестрации конвейеров данных, концепцию направленных ациклических графов и управление зависимостями задач;
- измерения и метрики качества данных, методы обработки пропусков, выбросов и временного дрейфа в потоках критической инфраструктуры;
- концепции управления происхождением данных, словари данных и основы Data Governance;
- принципы проектирования и документирования API для сервирования витрин данных потребителям;
- требования к информационной безопасности, защите критической информационной инфраструктуры и обезличиванию персональных данных в авиационных ИС;
- ландшафт отечественного программного обеспечения для задач хранения и обработки данных, стратегии импортозамещения ИТ-инфраструктуры;

- стандарты оформления архитектурной и проектной документации в соответствии с ЕСПД и ГОСТ при создании интеллектуальных систем;
- методологии кросс-функционального взаимодействия и формирования Service Level Agreement между командами Data Engineering и Data Science.

Уметь:

- уметь проектировать концептуальные схемы потоков данных при помощи нотаций UML и C4 в условиях гетерогенности авиационных источников;
- уметь разрабатывать контракты данных и спецификации схем при помощи форматов JSON Schema и Avro IDL с учетом требований к эволюции схем и обратной совместимости;
- уметь обосновывать выбор распределенных брокеров сообщений и паттернов ингестии при помощи матриц сравнения технологий в условиях пиковых нагрузок потоковой телеметрии;
- уметь проектировать архитектуру озер и хранилищ данных при помощи концепции Medallion с учетом стратегий партиционирования пространственно-временных рядов;
- уметь разрабатывать логические модели аналитических витрин при помощи ER-диаграмм и концепций звездной схемы для задач предиктивного ТОиР и мониторинга безопасности полетов;
- уметь формировать концептуальные графы оркестрации и спецификации SLA при помощи визуальных нотаций Apache Airflow в условиях жестких требований к задержкам интеллектуальных систем;
- уметь проектировать фреймворки обеспечения качества данных и прослеживаемости при помощи метрик полноты и непротиворечивости в контексте нормативов авиационной безопасности;
- уметь составлять архитектурные разделы Технических Заданий и спецификации API при помощи стандартов ГОСТ и OpenAPI для передачи спроектированных витрин данных командам Data Science и DevOps.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 5 з.е. (180 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
Контактная работа при проведении учебных занятий (всего):	32	32
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа	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 148 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	Введение в архитектуру больших данных и отраслевые стандарты аэронавигации Рассматриваемые вопросы: - парадигмы построения распределенных систем обработки больших данных и их применимость в авиации; - глобальные и отечественные стандарты обмена авиационной информацией, концепция SWIM и форматы AIXM, FIXM, WXXM; - ландшафт отечественного программного обеспечения и стратегии импортозамещения ИТ-инфраструктуры.
2	Моделирование потоков данных и архитектурные паттерны Рассматриваемые вопросы: - принципы концептуального, логического и физического моделирования потоков данных; - применение нотаций UML и модели C4 для проектирования ИТ-ландшафта; - специфика моделирования пространственно-временных авиационных данных.
3	Архитектура слоя ингестии и брокеры сообщений Рассматриваемые вопросы: - архитектура и сценарии применения распределенных брокеров сообщений;

№ п/п	Тематика лекционных занятий / краткое содержание
	- сравнительный анализ Apache Kafka, RabbitMQ и MQTT для потоковой телеметрии; - обеспечение отказоустойчивости и сглаживание пиковых нагрузок при приеме данных.
4	Форматы хранения и архитектура озер данных Рассматриваемые вопросы: - форматы сериализации и колоночного хранения полуструктурированных данных; - стратегии управления эволюцией схем и роль Schema Registry; - архитектурные паттерны организации озер данных и концепция многоуровневой архитектуры «Медальон».
5	Аналитические хранилища и СУБД для больших данных Рассматриваемые вопросы: - архитектура и отличительные особенности колоночных и транзакционных СУБД в ландшафте Big Data; - стратегии партиционирования, шардирования и индексирования массивов телеметрии; - концепции распределенных вычислений и in-memory обработки на уровне архитектурных возможностей.
6	Проектирование конвейеров данных и оркестрация Рассматриваемые вопросы: - парадигмы ETL и ELT, архитектурные различия и критерии выбора подхода; - принципы оркестрации конвейеров данных и концепция направленных ациклических графов; - управление зависимостями задач и формирование Service Level Agreement.
7	Управление качеством, происхождением данных и Data Governance Рассматриваемые вопросы: - измерения и метрики качества данных в контексте требований к безопасности полетов; - концепции управления происхождением данных и ведения словарей данных; - основы Data Governance и методологии кросс-функционального взаимодействия.
8	Сервисная архитектура, безопасность и документирование Рассматриваемые вопросы: - принципы проектирования и документирования API для сервирования витрин данных; - требования к информационной безопасности и защите критической информационной инфраструктуры; - стандарты оформления архитектурной и проектной документации в соответствии с ЕСПД и ГОСТ.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Аудит источников авиационных данных и концептуальное моделирование потоков Студент проводит классификацию гетерогенных источников телеметрии наземной инфраструктуры и бортовых систем воздушных судов. Выполняется анализ пропускной способности каналов связи и частоты генерации сигналов для определения паттернов поступления данных. На основе полученных метрик студент строит концептуальную схему потоков данных с использованием нотации C4. Итогом работы становится матрица характеристик источников и первичная диаграмма контекста системы.
2	Проектирование слоя ингестии и разработка контрактов данных Студент обосновывает выбор распределенного брокера сообщений для сглаживания пиковых нагрузок при приеме потоковой телеметрии ADS-B и датчиков перрона. Разрабатываются спецификации топиков и стратегии партиционирования очередей с учетом временных меток и идентификаторов рейсов. Студент формирует контракты данных в формате JSON Schema для зоны

№ п/п	Тематика практических занятий/краткое содержание
	приземления сырых данных. Результатом является схема слоя ингестии и набор валидационных правил для продюсеров сообщений.
3	Архитектурное проектирование озера данных и выбор форматов хранения Студент проектирует многоуровневую архитектуру хранения по паттерну «Медальон» с выделением сырой, очищенной и агрегированной зон. Выполняется сравнительный анализ форматов колоночного хранения для выбора оптимального решения под пространственно-временные ряды авиационной телеметрии. Разрабатывается стратегия физического партиционирования файлов по датам вылета и типам воздушных судов. Итогом служит архитектурная схема озера данных с обоснованием выбора технологий из реестра отечественного ПО.
4	Логическое моделирование аналитических витрин для предиктивного ТОиР Студент разрабатывает логическую модель витрины данных в виде схемы «звезда» для решения задач предиктивного обслуживания аэродромной инфраструктуры. Определяются таблицы фактов с метриками износа покрытий и таблицы измерений, содержащие справочники оборудования и метеорологических условий. Выполняется нормализация атрибутов и проектирование индексов для ускорения аналитических запросов. Результатом работы является ER-диаграмма витрины и спецификация DDL-конструкций для колоночной СУБД.
5	Концептуальное проектирование конвейеров трансформации данных Студент формирует матрицу алгоритмов очистки телеметрии от шумов датчиков и обработки пропущенных значений. Проектируется логика агрегации сырых потоков в интервальные метрики, необходимые для обучения моделей машинного обучения. Выполняется выбор вычислительных движков для пакетной и потоковой обработки с учетом ограничений по задержкам. Итогом становится концептуальная схема слоя трансформации и описание бизнес-логики преобразований.
6	Оркестрация процессов и формирование соглашений об уровне сервиса Студент проектирует направленный ациклический граф зависимостей задач для автоматизации запуска конвейеров данных. Определяются триггеры запуска, параметры повторных попыток при сбоях и механизмы оповещения об ошибках. Разрабатывается проект соглашения об уровне сервиса для гарантии своевременной доставки витрин данных команде Data Science. Результатом является визуальная схема оркестрации и спецификация SLA с метриками доступности системы.
7	Проектирование фреймворка обеспечения качества и прослеживаемости данных Студент разрабатывает систему метрик качества для автоматизированного контроля полноты и непротиворечивости авиационных данных. Проектируются механизмы отслеживания происхождения данных от источника до финальной витрины для соблюдения требований аудита. Формируются правила блокировки конвейера при обнаружении критических аномалий в телеметрии критической инфраструктуры. Итогом работы является паспорт качества данных и схема lineage для сквозного мониторинга.
8	Формирование архитектурного манифеста и технического задания Студент консолидирует все разработанные артефакты в единый альбом архитектурных решений платформы больших данных. Выполняется подготовка фрагмента технического задания для команды разработки с описанием требований к развертыванию инфраструктуры. Формируются спецификации API в формате OpenAPI для предоставления доступа к витринам данных внешним потребителям. Результатом является итоговый архитектурный манифест и пакет проектной документации, готовый к передаче в производство.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение рекомендованной литературы.
2	Подготовка к промежуточной аттестации.

3	Подготовка к текущему контролю.
---	---------------------------------

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

№ п/п	Библиографическое описание	Место доступа
1	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 188 с. — ISBN 978-5-507-47346-5. — Текст : электронный	Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362318 (дата обращения: 25.06.2026)
2	Параскевов, А. В. Большие данные : учебник / А. В. Параскевов, А. Э. Сергеев. — Вологда : Инфра-Инженерия, 2024. — 148 с. — ISBN 978-5-9729-2120-1. — Текст : электронный	Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/427811 (дата обращения: 25.06.2026)
3	Талия, Д. Большие данные. Современные фреймворки и разработка приложений : руководство / Д. Талия, П. Трунфио, Ф. Мароццо ; пер. с англ. А. В. Логунова. — Москва : ДМК Пресс, 2025. — 272 с. — ISBN 978-5-93700-358-4. — Текст : электронный	Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/514895 (дата обращения: 25.06.2026)
4	Толстобров, А. П. Управление данными : учебник для вузов / А. П. Толстобров. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 272 с. — (Высшее образование). — ISBN 978-5-534-14162-7. — Текст : электронный	Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/588522 (дата обращения: 25.06.2026)
5	Орешков, В. И. Хранилища данных и OLAP-технологии : учебное пособие / В. И. Орешков. — Рязань : РГРТУ, 2017. — 64 с. — Текст : электронный	Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/167981 (дата обращения: 25.06.2026)
6	Ланских, Ю. В. Введение в большие данные : учебное пособие / Ю. В. Ланских, В. Г. Ланских, К. В. Родионов. — Киров : ВятГУ, 2023. — 172 с. — Текст : электронный	Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/408566 (дата обращения: 25.06.2026)

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

Официальная документация колоночной СУБД ClickHouse – <https://clickhouse.com/docs/ru>

Официальная русскоязычная документация СУБД Postgres Pro (Регистр ПО РФ) – <https://postgrespro.ru/docs>

Официальная документация распределенного брокера сообщений Apache Kafka – <https://kafka.apache.org/documentation/>

Официальная документация платформы оркестрации конвейеров данных Apache Airflow – <https://airflow.apache.org/docs/>

Спецификация OpenAPI для проектирования и документирования REST API – <https://swagger.io/specification/>

Концепция SWIM и стандарты обмена авиационными данными (AIXM, FIXM, WXXM) // Eurocontrol / ICAO Official Documentation – <https://www.eurocontrol.int/swim>

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

Операционные системы – Astra Linux Special Edition, РЕД ОС.

Офисные пакеты – Р7-Офис, МойОфис.

Инструменты моделирования и проектирования – PlantUML, Draw.io, StarUML.

Инструменты спецификации данных и API – JSON Schema, Apache Avro IDL, Swagger Editor.

Целевой технологический стек ИТ-ландшафта – Apache Kafka, RabbitMQ, MQTT, Apache Parquet, Apache Avro, ClickHouse, Apache Greenplum, Postgres Pro, Apache Hadoop, Apache Airflow, Apache Atlas.

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

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

Для практических занятий – наличие персональных компьютеров вычислительного класса.

9. Форма промежуточной аттестации:

Зачет в 1 семестре.

10. Оценочные материалы.

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

Авторы:

заведующий кафедрой, доцент, к.н.
кафедры «Цифровые технологии
управления транспортными
процессами»

В.Е. Нутович

Согласовано:

Заместитель директора академии

Д.Е. Гончаров

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

Н.А. Андриянова