

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 РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Технологии хранения больших данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте (Российско-Китайская программа)

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 РУТ (МИИТ)
ID подписи: 937226
Подписал: руководитель образовательной программы
Проневич Ольга Борисовна
Дата: 15.06.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины «Технологии хранения больших данных» является теоретическая и практическая подготовка студентов к работе с большими данными.

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-3 - Способен использовать современные информационные технологии и программно-аппаратные средства, в том числе отечественного производства, при решении задач профессиональной деятельности;

ПК-1 - Способен анализировать большие данные с использованием существующей в организации методологической и технологической инфраструктуры.

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

Знать:

методы анализа больших данных, интегрируемых в корпоративные информационные системы,

методы хранения и обработки больших данных,

IT-сервисы и программные обеспечения, необходимые для работы с большими данными

Уметь:

находить данные для анализа в открытых источниках,

загружать данные, проводить предварительную чистку данных для анализа,

проводить анализ данных в зависимости от их типа

проводить установку Apache Hadoop

настраивать в минимальных требованиях Apache Hadoop

тестировать Apache Hadoop

Владеть:

Инструментами и командами языка R для анализа данных
Знаниями об определении типа диаграмм
Статистическими параметрами, коэффициентами и тестами
Инструментами MCA и CA анализа
Тремя режимами установки Apache Hadoop
Системой установки Spark

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 3 з.е. (108 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):	64	64
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	32	32

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 44 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	<p>Тема 1. Проекты в области Больших Данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - CRISP-DM. Этапы. - Роли в проекте: Data Engineer, Data Analyst, Data Scientist.
2	<p>Тема 2. Многомерный статистический анализ.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Коэффициент корреляции. - Факторный анализ.
3	<p>Тема 3. Методы многомерного статистического анализа.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Оценка качества модели. - Многомерное шкалирование.
4	<p>Тема 4. Методы многомерного статистического анализа.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Дискриминантный анализ. - Методы контроля качества.
5	<p>Тема 5. Методы многомерного статистического анализа.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Иерархический кластерный анализ. - Метод K-means.
6	<p>Тема 6. Технологии хранения и обработки больших данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - OLTP и OLAP системы. - DataWareHouse, DataLake и LakeHouse. - Модели Кимбалла и Инмона. - Схема Звезда и Снежинка. - Архитектура слоев хранения данных. - Методы загрузки данных: ETL и ELT.
7	<p>Тема 7. Проектирование MPP-систем.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - DAMA DMBоК. - Data Vault и Anchor Modeling. - Вертикальное и горизонтальное масштабирование. - Вертикальное шардирование (партиционирование). - Горизонтальное шардирование. - MPP – архитектура (Massively Parallel Processing). - Пакетная и потоковая обработка данных. - Лямбда – архитектура.
8	<p>Тема 8. Хранение неструктурированных данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - ЦОД – TIER – RAID. - Проблема хранения неструктурированных данных: ACID и BASE. - Теорема CAP. - Реляционные СУБД и NoSQL. - Модели данных NoSQL.

№ п/п	Тематика лекционных занятий / краткое содержание
9	Тема 9. Парадигма MapReduce, Hadoop, HDFS Рассматриваемые вопросы: - Архитектура Hadoop. - HDFS. - Компоненты и планировщики YARN. - Концепция Map Reduce.
10	Тема 10. Apache Spark. Рассматриваемые вопросы: - Загрузка данных в Hadoop. - Spark Context. - Framework Apache Spark. - RDD. - Spark Streaming.
11	Тема 11. SQL поверх Hadoop. Рассматриваемые вопросы: - Архитектура Hive. - Развертывание MetaStore. - Trino.
12	Тема 12. Экосистема Hadoop. Рассматриваемые вопросы: - Apache Kafka. - Топики и партиции Kafka. - Apache ZooKeeper.
13	Тема 13. Технологии и методы анализа Big Data. Рассматриваемые вопросы: - Статистический анализ. - Метод смещения и интеграции данных. - Машинное обучение и Нейросети. - Data Mining. - Стратегия краудсорсинга. - Метод предиктивной аналитики. - Технология имитационного моделирования. - Визуализация аналитических данных – BI системы.
14	Тема 14. Языки программирования для Big Data. Рассматриваемые вопросы: - Архитектура Фон Неймана. - Закон Мура. - R, Python, SQL, Java, Scala, Julia, Go, C++.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. Методы многомерного статистического анализа. Рассматриваемые вопросы: - Практическое занятие по факторному анализу.

№ п/п	Тематика практических занятий/краткое содержание
2	Тема 2. Методы многомерного статистического анализа. Рассматриваемые вопросы: - Анализ данных с помощью Python. - Библиотеки Python: pandas, numpy, matplotlib, seaborn, scikit-learn
3	Тема 3. Методы многомерного статистического анализа. Рассматриваемые вопросы: - Анализ данных с помощью SQL.
4	Тема 4. Методы многомерного статистического анализа. Рассматриваемые вопросы: - Исследование датасета и кластеризация данных. - A/B анализ.
5	Тема 5. Технологии хранения и обработки больших данных. Рассматриваемые вопросы: - Работа с ETL сервисами.
6	Тема 6. Хранение неструктурированных данных. Рассматриваемые вопросы: - MongoDB. - Колочное хранение данных.
7	Тема 7. Парадигма MapReduce, Hadoop, HDFS. Рассматриваемые вопросы: - Word Count с использованием Map Reduce. - Узлы кластера Hadoop.
8	Тема 8. Apache Spark. Рассматриваемые вопросы: - Анализ данных с помощью Apache Spark. - Работа с PySpark.
9	Тема 9. SQL поверх Hadoop. Рассматриваемые вопросы: - HiveQL. - Внутренние и внешние таблицы Hive.
10	Тема 10. Экосистема Hadoop Рассматриваемые вопросы: Компоненты кластера Apache Kafka. Коэффициент репликации Kafka.
11	Тема 11. Технологии и методы анализа Big Data Рассматриваемые вопросы: - Методы разметки данных. - Визуализация данных и работа с BI-инструментарием.
12	Тема 12. Языки программирования для Big Data. Рассматриваемые вопросы: - Сравнение производительности языков программирования.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение и конспектирование литературы по дисциплине.

№ п/п	Вид самостоятельной работы
2	Подготовка к текущему контролю
3	Подготовка к практическим занятиям
4	Подготовка к промежуточной аттестации.
5	Выполнение курсовой работы.
6	Подготовка к промежуточной аттестации.
7	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

Анализ, обработка, визуализация пассажиропотоков Московского транспортного узла (варианты по видам транспорта и направлениям)

1. Применение R для анализа данных
2. Методы анализа больших данных
3. Факторный анализ больших данных
4. Аналитика Big Data — реалии и перспективы в России и мире.
5. Технологии и методы анализа, которые используются для анализа Big Data
6. Источники данных о пассажиропотоках: турникеты метро, валидаторы наземного транспорта, GPS-трекеры, данные сотовых операторов, системы видеонаблюдения.
7. Очистка и предобработка транспортных данных: устранение пропусков, фильтрация аномальных значений, агрегация по часам и направлениям.
8. Визуализация загруженности маршрутов: тепловые карты, хороплеты, анимированные графики движения, Sankey-диаграммы пересадок.
9. Анализ пассажиропотоков по видам транспорта: метро, МЦК, МЦД, автобус, электробус, такси — сравнение суточной динамики.
10. Прогнозирование пассажиропотоков с использованием временных рядов (ARIMA, Prophet) и методов машинного обучения.
11. Разработка дашборда для мониторинга пассажиропотоков в реальном времени (R Shiny, Tableau, Power BI, Yandex DataLens).

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

№ п/п	Библиографическое описание	Место доступа
1	<p>Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060-618-6</p>	<p>https://e.lanbook.com/book/107901</p>
2	<p>Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7</p>	<p>https://e.lanbook.com/book/69955</p>
3	<p>Габдуллин, Н. М. Развитие человеческого капитала и цифровой экономики в регионах России: факторный и кластерный анализ : монография / Н. М. Габдуллин. — Казань : КФУ, 2019. — 268 с. — ISBN 978-5-00130-291-9</p>	<p>https://e.lanbook.com/book/173018</p>
4	<p>Гласснер, Э. Глубокое обучение без математики. Том 2. Практика : руководство / Э. Гласснер ; перевод с английского В. А. Яроцкого. — Москва : ДМК Пресс, 2020. — 610 с. — ISBN 978-5-97060-767-1</p>	<p>https://e.lanbook.com/book/131710</p>
5	<p>Гульятеева, Т. А. Методы статистического обучения в задачах регрессии и классификации :</p>	<p>https://e.lanbook.com/book/118291</p>

	монография / Т. А. Гульятеева, А. А. Попов, А. С. Саутин. — Новосибирск : НГТУ, 2016. — 323 с. — ISBN 978-5-7782-2817-7	
6	Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук ; перевод с английского А. Б. Огурцова. — Москва : ДМК Пресс, 2018. — 250 с. — ISBN 978-5-97060-508-0	https://e.lanbook.com/book/97353
7	Шалев-Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев-Шварц, Бен-Давид Ш. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5	https://e.lanbook.com/book/131686
8	Изучаем Spark: молниеносный анализ данных / Х. Карау, Э. Конвински, П. Венделл, М. Захария. — Москва : ДМК Пресс, 2015. — 304 с. — ISBN 978-5-97060-323-9	https://e.lanbook.com/book/90118?ysclid=1wkcbpunxi226114464

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

<https://journal.tinkoff.ru/short/ai-for-all/> - база данных нейронных сетей

<https://vc.ru/services/916617-luchshie-neyroseti-bolshaya-podborka-iz-top-200-ii-generatorov-po-kategoriyam> - база данных нейронных сетей

<https://github.com/abalmumcu/bert-rest-api> - профессиональная платформа для командой работы над проектов (нейронная сеть bert)

<http://library.miit.ru/> - электронно-библиотечная система Научно-технической библиотеки МИИТ

<https://proglib.io/p/raspoznavanie-obektov-s-pomoshchyu-yolo-v3-na-tensorflow-2-0-2020-11-08> - профессиональная библиотека программистов

https://yandex.cloud/ru/blog/posts/2022/12/andrey-berger-and-yandex-cloud?utm_referrer=https%3A%2F%2Fyandex.ru%2F – библиотека профессиональных статей разработчиков Яндекс

<https://yandex.cloud/ru/blog> - библиотека профессиональных статей разработчиков Яндекс

<https://tproger.ru/translations/opencv-python-guide> - библиотека основных команд OpenCV

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

Windows 7 и выше

Microsoft Office 2010

R consoler

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

Компьютер преподавателя

Компьютеры студентов

Монитор

Проектор Optoma W340UST

Экран для проектора

Маркерная доска

9. Форма промежуточной аттестации:

Зачет в 6 семестре.

Курсовая работа в 6 семестре.

10. Оценочные материалы.

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

Авторы:

заместитель директора

Б.В. Игольников

руководитель образовательной
программы

О.Б. Проневич

Согласовано:

Директор

Д.В. Паринов

Руководитель образовательной
программы

О.Б. Проневич

Председатель учебно-методической
комиссии

Д.В. Паринов