

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



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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Технологии больших данных

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Компьютерные сети и технологии

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: 4196
Подписал: заведующий кафедрой Желенков Борис
Владимирович
Дата: 22.10.2025

1. Общие сведения о дисциплине (модуле).

Целями освоения дисциплины (модуля) являются:

- изучение современных методов организации больших баз данных;
- изучение программных средств, используемых для организации хранения и обработки больших баз данных;
- изучение методов и программных средств для анализа больших баз данных.

Задачами освоения дисциплины (модуля) являются:

- разработка структур больших баз данных;
- написание запросов, работающих с большими базами данных;
- разработка приложений, использующих большие базы данных.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-8 - Способен осуществлять эффективное управление разработкой программных средств и проектов.;

ПК-5 - Владение методами и алгоритмами решения задач обработки данных;

ПК-7 - Применение перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий;

ПК-8 - Понимание подходов к верификации моделей программного обеспечения.

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

Знать:

- основные понятия больших баз данных,
- технологии построения распределенных баз данных,
- языки и алгоритмы обработки больших данных,
- методы верификации распределенных баз данных.

Уметь:

- описывать таблицы и материализованные представления,
- разрабатывать запросы на языке CQL,
- разрабатывать программы на языке Python с использованием Spark,
- разрабатывать модели анализа данных.

Владеть:

- языками описания схем баз данных,
- языками запросов к большим базам данных,
- средами разработки и отладки программ работы с большими данными,
- средствами проектирования моделей данных.

3. Объем дисциплины (модуля).**3.1. Общая трудоемкость дисциплины (модуля).**

Общая трудоемкость дисциплины (модуля) составляет 6 з.е. (216 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 №3
Контактная работа при проведении учебных занятий (всего):	48	48
В том числе:		
Занятия лекционного типа	32	32
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 168 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	Большие данные, анализ данных и наука о данных Рассматриваемые вопросы: - основные понятия - источники больших данных, - анализ данных, - типы анализируемых данных, - процесс обработки данных.
2	Система управления базами данных Cassandra Рассматриваемые вопросы: - установка системы, - работа с системой из командной строки, - модель данных, - пространство ключей, - таблицы и материализованные представления, - операции обновления данных, - выполнение запросов.
3	Распределение данных в системе Cassandra Рассматриваемые вопросы: - параллельные архитектуры, - масштабирование баз данных, - распределение данных в Cassandra, - теорема CAP, - обнаружение отказов, - репликация данных
4	Установка систем Spyder и Spark Рассматриваемые вопросы: - установка Spyder, - установка pip, - установка Spark.
5	Система анализа Spark Рассматриваемые вопросы: - использование Spark в Spyder, - наборы RDD, - выполнение программы, - Spark и Cassandra, - работа с набором данных в Spyder
6	Spark SQL Рассматриваемые вопросы: - выполнение запросов,
7	Spark SQL(продолжение) Рассматриваемые вопросы: - соединение таблиц,
8	Spark SQL(продолжение) Рассматриваемые вопросы: - функции агрегирования.
9	Технология MapReduce Рассматриваемые вопросы: - процесс вычислений, - функция map, - функция reduce,

№ п/п	Тематика лекционных занятий / краткое содержание
10	Технология MapReduce(продолжение) Рассматриваемые вопросы: - умножение матрицы на вектор и на матрицу
11	Технология MapReduce(продолжение) Рассматриваемые вопросы: - реализация операций реляционной алгебры
12	Поиск похожих документов Рассматриваемые вопросы: - сходство множеств, - представление документа в виде множеств, - шинглы, - матричное представление множеств,
13	Поиск похожих документов(продолжение) Рассматриваемые вопросы: - минхэш-сигнатуры, - алгоритм определения похожих документов, - метрики.
14	Поиск похожих документов в Spark Рассматриваемые вопросы: - сравнение двух множеств документов, - сравнение документов с образцом.
15	Линейная регрессия Рассматриваемые вопросы: - машинное обучение, - оценка модели, - библиотека функций машинного обучения, - линейная регрессия. - обучающий и тестовый наборы, - предсказание значений.
16	Линейная регрессия в Spark Рассматриваемые вопросы: - построение модели, - график простой линейной регрессии.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Разработка базы данных в системе Cassandra. В результате выполнения лабораторной работы студент получает навык проектирования базы данных в NoSQL системе управления базами данных Cassandra.
2	Обработка запросов в системе Cassandra. В результате выполнения лабораторной работы студент получает навык выполнения запросов в NoSQL системе управления базами данных Cassandra.
3	Анализ данных в системе Spark. Выполнение запросов. В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.

№ п/п	Наименование лабораторных работ / краткое содержание
4	Анализ данных в системе Spark. Выполнение запросов, использующих несколько таблиц. В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.
5	Анализ данных в системе Spark. Выполнение запросов, использующих функции агрегирования. В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.
6	Анализ данных в системе Spark. Выявление похожих объектов. В результате выполнения лабораторной работы студент получает навык сравнения документов в системе Spark и выявления похожих документов.
7	Анализ данных в системе Spark. Линейная регрессия. В результате выполнения лабораторной работы студент получает навык использования алгоритма линейной регрессии для прогнозирования значений данных.
8	Анализ данных в системе Spark. Визуализация простой линейной регрессии. В результате выполнения лабораторной работы студент получает навык использования алгоритма линейной регрессии для прогнозирования значений данных.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение документации по системе Cassandra
2	Подготовка к лабораторным работам
3	Изучение учебной литературы из приведенных источников
4	Анализ и дополнительная проработка лекционного материала
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Мамедли Р. Э., Казиахмедов Т. Б. Большие данные и NoSQL базы данных: Учебное пособие для вузов. - Издательство "Лань". 2024. 92 с. ISBN: 978-5-507-49873-4	https://reader.lanbook.com/book/434051#1 (дата обращения: 04.12.2024). - Текст: электронный.
2	Макшанов А. В., Журавлев А. Е., Тындыкаръ Л. Н. Большие данные. Big Data: Учебник для вузов. - Издательство "Лань". 2024. 188 с. ISBN: 978-5-507-47346-5	https://reader.lanbook.com/book/362318 (дата обращения: 04.12.2024). - Текст: электронный.

3	Ланских Ю. В., Ланских В. Г., Родионов К. В. Введение в большие данные: учеб. пособие. - Вятский государственный университет. 2023. 172 с.	https://reader.lanbook.com/book/408566 (дата обращения: 04.12.2024). - Текст: электронный.
---	--	---

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

- Научная электронная библиотека (<http://elibrary.ru>)
- Электронно-библиотечная система Научно-технической библиотеки МИИТ (<http://library.miit.ru>)
- Википедия (<https://ru.wikipedia.org>)
- Материалы по информационным технологиям (www.citforum.ru)
- Сайт системы Cassandra (<http://cassandra.apache.org>)

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

- Microsoft Windows
- Microsoft Office
- Интернет-браузер (Yandex и др.)
- Язык программирования Python,
- Система баз данных Apache Cassandra (лицензия – свободно-распространяемое ПО с открытым кодом)

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

Учебная аудитория для проведения учебных занятий (занятий лекционного типа, практических занятий, лабораторных работ, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации):

- компьютер преподавателя, проектор, экран проекционный, рабочие станции студентов, маркерная доска.

Аудитория подключена к сети «Интернет»

9. Форма промежуточной аттестации:

Зачет в 3 семестре.

10. Оценочные материалы.

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

Авторы:

доцент, доцент, к.н. кафедры
«Вычислительные системы, сети и
информационная безопасность»

М.А. Давыдовский

Согласовано:

Заведующий кафедрой ВССиИБ

Б.В. Желенков

Председатель учебно-методической
комиссии

Н.А. Андриянова