

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: 22.05.2024

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.

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

Знать:

- основные понятия больших баз данных и технологии построения распределенных баз данных.

Уметь:

- разрабатывать схемы и запросы для больших баз данных.

Владеть:

- языками описания и запросов к большим базам данных, средствами проектирования моделей данных.

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 №2
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 132 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:

№ п/п	Тематика лекционных занятий / краткое содержание
	<ul style="list-style-type: none"> - установка системы, - работа с системой из командной строки, - модель данных, - пространство ключей, - таблицы и материализованные представления, - операции обновления данных, - выполнение запросов.
3	<p>Распределение данных в системе Cassandra</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - параллельные архитектуры, - масштабирование баз данных, - распределение данных в Cassandra, - теорема CAP, - обнаружение отказов, - репликация данных
4	<p>Установка систем Spyder и Spark</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - установка Spyder, - установка pip, - установка Spark.
5	<p>Система анализа Spark</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - использование Spark в Spyder, - наборы RDD, - выполнение программы, - Spark и Cassandra, - работа с набором данных в Spyder
6	<p>Spark SQL</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - выполнение запросов,
7	<p>Spark SQL(продолжение)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - соединение таблиц,
8	<p>Spark SQL(продолжение)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - функции агрегирования.
9	<p>Технология MapReduce</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - процесс вычислений, - функция map, - функция reduce,
10	<p>Технология MapReduce(продолжение)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - умножение матрицы на вектор и на матрицу
11	<p>Технология MapReduce(продолжение)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - реализация операций реляционной алгебры
12	<p>Поиск похожих документов</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - сходство множеств,

№ п/п	Тематика лекционных занятий / краткое содержание
	<ul style="list-style-type: none"> - представление документа в виде множеств, - шинглы, - матричное представление множеств,
13	<p>Поиск похожих документов(продолжение)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - минхэш-сигнатуры, - алгоритм определения похожих документов, - метрики.
14	<p>Поиск похожих документов в Spark</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - сравнение двух множеств документов, - сравнение документов с образцом.
15	<p>Линейная регрессия</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - машинное обучение, - оценка модели, - библиотека функций машинного обучения, - линейная регрессия. - обучающий и тестовый наборы, - предсказание значений.
16	<p>Линейная регрессия в Spark</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - построение модели, - график простой линейной регрессии.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	<p>Разработка базы данных в системе Cassandra.</p> <p>В результате выполнения лабораторной работы студент получает навык проектирования базы данных в NoSQL системе управления базами данных Cassandra.</p>
2	<p>Обработка запросов в системе Cassandra.</p> <p>В результате выполнения лабораторной работы студент получает навык выполнения запросов в NoSQL системе управления базами данных Cassandra.</p>
3	<p>Анализ данных в системе Spark. Выполнение запросов.</p> <p>В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.</p>
4	<p>Анализ данных в системе Spark. Выполнение запросов, использующих несколько таблиц.</p> <p>В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.</p>
5	<p>Анализ данных в системе Spark. Выполнение запросов, использующих функции агрегирования.</p> <p>В результате выполнения лабораторной работы студент получает навык выполнения запросов в системе Spark к данным, хранящимся в базе данных системы Cassandra.</p>
6	<p>Анализ данных в системе Spark. Выявление похожих объектов.</p>

№ п/п	Наименование лабораторных работ / краткое содержание
	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	Царькова Н.И., Смолянов А.С. BigData. Развитие, анализ и технологии. Актуальные проблемы гуманитарных и естественных наук. 2016. № 7-1. - с.86-95	https://elibrary.ru/item.asp?id=26365870 (дата обращения: 03.04.2024). - Текст: электронный.
2	Порохненко Ю.С., Полежаев П.Н. Современные технологии NoSQL для реализации баз данных. Актуальные направления научных исследований XXI века: теория и практика. 2017. Т. 5. № 9 (35). – с.25-30	https://elibrary.ru/item.asp?id=32246848 (дата обращения: 03.04.2024). - Текст: электронный.
3	Золотов В.А., Семенов В.А. Сравнительный анализ NoSQL баз данных. Труды Института системного программирования РАН. 2013. Т. 25. – с.131-166	https://elibrary.ru/item.asp?id=21134920 (дата обращения: 03.04.2024). - Текст: электронный.
4	Григорьев Ю.А. Анализ свойств баз данных NoSQL. Информатика и системы управления. 2013. № 2 (36). С. 003-013	https://elibrary.ru/item.asp?id=19060114 (дата обращения: 03.04.2024). - Текст: электронный.

5	Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. Анализ больших наборов данных. - М.: ДМК Пресс, 2016. - 498 с.	https://e.lanbook.com/book/93571 (дата обращения: 03.04.2024). - 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.miiit.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).

- Язык программирования Python,
- Система баз данных Apache Cassandra (лицензия – свободно-распространяемое ПО с открытым кодом)

При организации обучения по дисциплине (модулю) с применением электронного обучения и дистанционных образовательных технологий необходим доступ каждого студента к информационным ресурсам – библиотечному фонду Университета, сетевым ресурсам и информационно-телекоммуникационной сети «Интернет».

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий может понадобиться наличие следующего программного обеспечения (или их аналогов): ОС Windows, Microsoft Office, Интернет-браузер, Microsoft Teams и т.д.

В образовательном процессе, при проведении занятий с применением электронного обучения и дистанционных образовательных технологий, могут применяться следующие средства коммуникаций: ЭИОС РУТ(МИИТ), Microsoft Teams, электронная почта, скайп, Zoom, WhatsApp и т.п.

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

- Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций .

Проектор для вывода изображения на экран для студентов, акустическая система, место для преподавателя оснащенное компьютером. Аудитория подключена к интернету МИИТ.

- Учебная аудитория для проведения занятий лекционного типа, практических занятий, лабораторных работ.

Аудиовизуальное оборудование для аудитории, АРМ управляющий, проектор, экран, персональные компьютеры, мониторы, принтер, доска учебная. Аудитория подключена к интернету МИИТ.

- В случае проведении занятий с применением электронного обучения и дистанционных образовательных технологий необходимо наличие компьютерной техники, для организации коллективных и индивидуальных форм общения педагогических работников со студентами, посредством используемых средств коммуникации.

Допускается замена оборудования его виртуальными аналогами.

9. Форма промежуточной аттестации:

Зачет во 2 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