

**Федеральное государственное образовательное бюджетное учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ
ФЕДЕРАЦИИ»**

(Финансовый университет)

Новороссийский филиал

Кафедра «Информатика, математика и общегуманитарные науки»

УТВЕРЖДАЮ

Директор филиала


Н. Н. Сейфидзе

« 31 »

марта 2022



Б.Б. Зайковский

ОБРАБОТКА СТАТИСТИЧЕСКИХ ПОТОКОВЫХ БОЛЬШИХ ДАННЫХ

Рабочая программа

дисциплины

для студентов, обучающихся по направлению подготовки:

09.04.03 Прикладная информатика, направленность программы магистратуры:

Обработка больших данных и разработка интеллектуальных приложений

Рекомендовано Ученым советом Новороссийского филиала Финансового университета

протокол № 45 от 31 марта 2022 г.

Одобрено кафедрой «Информатика, математика и общегуманитарные

науки» протокол № 8 от 24 марта 2022 г.

Новороссийск 2022

СОДЕРЖАНИЕ

1. Наименование дисциплины.....	3
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3. Место дисциплины в структуре образовательной программы.....	5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	5
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	6
5.1. Содержание дисциплины.....	6
5.2. Учебно–тематический план.....	7
5.3. Содержание семинаров, практических занятий.....	7
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	9
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	9
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	9
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	10
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	16
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	17
10. Методические указания для обучающихся по освоению дисциплины	18
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	18
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	19

1. Наименование дисциплины

«Обработка статистических потоковых больших данных».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Таблица 1

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
УК-3	Способность определять и реализовывать приоритеты собственной деятельности в соответствии с важностью задач, методы повышения ее эффективности	1. Объективно оценивает свои возможности и требования различных социальных ситуаций. Принимает решения в соответствии с данной оценкой и требованиями.	Знать: принципы объединения данных из различных источников информации для последующего их анализа и обработки. Уметь: формулировать и решать задачи обработки больших данных; уметь обращаться к внешним ресурсам для получения данных.
		2. Актуализирует свой личностный потенциал, внутренние источники роста и развития собственной деятельности.	Знать: области применения технологий обработки больших данных, принципы построения систем массово-параллельной обработки данных. Уметь: формулировать и решать задачи обработки больших данных; уметь обращаться к внешним ресурсам для получения данных.
		3. Определяет приоритеты собственной деятельности в соответствии с важностью задач.	Знать: основные принципы обработки статистических потоковых больших данных. Уметь: использовать специализированное программное обеспечение для решения задач распределенной обработки больших объемов информации.

		4. Определяет и демонстрирует методы повышения эффективности собственной деятельности.	<p>Знать: области применения технологий обработки больших данных, принципы построения систем массово-параллельной обработки данных.</p> <p>Уметь: проектировать структуры для обработки больших объемов информации при решении специализированных задач в различных прикладных областях.</p>
ПКН-6	Способность анализировать предметную область и исследовать ИТ-рынок для решения задач информатизации, готовить аналитические обзоры с обоснованными выводами и рекомендациями	1. Анализирует предметную область и ИТ-рынок для решения задач информатизации.	<p>Знать: принципы объединения данных из различных источников информации для последующего их анализа и обработки.</p> <p>Уметь: формулировать и решать задачи обработки больших данных; уметь обращаться к внешним ресурсам для получения данных.</p>
		2. Готовит аналитические обзоры с обоснованными выводами и рекомендациями.	<p>Знать: основные принципы обработки статистических потоковых больших данных.</p> <p>Уметь: готовить аналитические обзоры с обоснованными выводами использования специализированное программное обеспечение для решения задач распределенной обработки больших объемов информации.</p>

3. Место дисциплины в структуре образовательной программы

Дисциплина «Обработка статистических потоковых больших данных» является дисциплиной Модуля общепрофессиональных дисциплин направления подготовки 09.04.03-Прикладная информатика, направленности программы магистратуры: Обработка больших данных и разработка интеллектуальных приложений, Управление большими данными.

Изучение дисциплины «Обработка статистических потоковых больших данных» базируется на знаниях, полученных в рамках дисциплин: «Нереляционные базы данных», «Проектирование информационных систем», «Современные компьютерные технологии машинного обучения».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

*Обработка больших данных и разработка интеллектуальных приложений, /
Управление большими данными (Институт онлайн-образования)*

Заочная форма обучения

Таблица 2

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Модуль 8 / 5 (в часах)
Общая трудоёмкость дисциплины	4/144	144
<i>Контактная работа – Аудиторные занятия</i>	<i>16</i>	<i>16</i>
<i>Лекции</i>	<i>4</i>	<i>4</i>
<i>Семинары, практические занятия</i>	<i>12</i>	<i>12</i>
Самостоятельная работа	128	128
Вид текущего контроля		Контрольная работа
Вид промежуточной аттестации		экзамен / зачет

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Подходы к обработке больших данных

Основные понятия. Подходы к обработке больших данных. Экосистема Hadoop. Структура и принципы построения кластерных систем. Развертывание кластера. Прием данных для пакетной и интерактивной обработки: прием данных из облака или локальных данных. Форматы файлов. Работа с текстовыми файлами. Файлы JSON. Значения, разделенные запятыми, и значения, разделенные табуляцией. SequenceFile. Объектные файлы. Форматы Hadoop для ввода и вывода. Сжатие файлов.

Тема 2. Обработка данных в Apache Spark

Управление Hadoop и Spark: создание и настройка кластера. Загрузка Spark. Введение в командные оболочки Spark. Введение в основные понятия Spark. Автономные приложения. Инициализация SparkContext. Отладка заданий Hadoop и Spark. Прием данных в Apache Spark. Программирование операций с RDD. Основы RDD. Создание RDD. Операции с RDD. Преобразования и действия. Отложенные вычисления. Передача функций в Spark.

Тема 3. SparkSQL

Работа с парами ключ/значение в Spark. Создание наборов пар. Преобразование наборов пар. Агрегирование, группировка, соединение, группировка. Управление распределением данных. Аккумуляторы. Широковещательные переменные. Работа с разделами по отдельности. Настройка Spark с помощью SparkConf. Включение SparkSQL в приложения. Использование SparkSQL в приложениях. Инициализация SparkSQL. Набор данных SchemaRDD. Кэширование. Загрузка и сохранение данных. Производительность SparkSQL.

Тема 4. Обработка потоковых данных

Архитектуры и абстракции потоковой обработки больших данных. Spark Streaming. Преобразования без сохранения состояния. Преобразования с сохранением состояния. Операции вывода. Источники исходных данных. Основные и дополнительные источники данных. Отказоустойчивость рабочих узлов и приемников. Проблемы производительности. Интервал пакетирования и протяженность окна. Степень параллелизма. Сборка мусора и использование памяти.

5.2. Учебно–тематический план

Таблица 3

№ п/п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах				Самостоятельная работа	Формы текущего контроля успеваемости
		Всего	Контактная работа- Аудиторная работа				
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Подходы к обработке больших данных	34	2	2	-	32	Самостоятельные работы. Участие в решении задач на практических занятиях. Обсуждение решенных задач.
2.	Обработка данных в Apache Spark	38	6	2	4	32	
3.	SparkSQL	36	4	-	4	32	
4.	Обработка потоковых данных	36	4	-	4	32	
	В целом по дисциплине	144	16	4	12	128	Согласно учебному плану: контрольная работа
	Итого в %		11	25	75	89	

5.3. Содержание семинаров, практических занятий

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Подходы к обработке больших данных	Развертывание кластера. Прием данных для пакетной и интерактивной обработки: прием данных из облака или локальных данных. Работа с текстовыми файлами. Файлы JSON. Значения, разделенные запятыми, и значения, разделенные табуляцией. SequenceFile. Объектные файлы.	-работа с текстом лекции; -изучение рекомендованных к занятию литературных источников;

	<p>Форматы Hadoop для ввода и вывода. Сжатие файлов.</p> <p><i>Рекомендуемые источники:</i> <i>Основная литература – 8.[1,2], 9.[24]</i> <i>Дополнительная литература - 8.[4,5], 9.[24]</i></p>	<p>-подготовка к семинарским и практическим занятиям; - выполнение домашних заданий</p>
Обработка данных в Apache Spark	<p>Загрузка Spark. Инициализация SparkContext. Прием данных в Apache Spark. Программирование операций с RDD. Основы RDD. Создание RDD. Операции с RDD. Преобразования и действия. Отложенные вычисления. Передача функций в Spark.</p> <p><i>Рекомендуемые источники:</i> <i>Основная литература – 8.[2,3], 9.[24]</i> <i>Дополнительная литература - 8.[5], 9.[24]</i></p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарским и практическим занятиям; -выполнение домашних заданий</p>
SparkSQL	<p>Работа с парами ключ/значение в Spark. Агрегирование, группировка, соединение, группировка. Аккумуляторы. Широковещательные переменные. Настройка Spark с помощью SparkConf. Использование SparkSQL в приложениях. Инициализация SparkSQL. Набор данных SchemaRDD. Кэширование. Загрузка и сохранение данных.</p> <p><i>Рекомендуемые источники:</i> <i>Основная литература – 8.[1,2,3], 9.[24]</i> <i>Дополнительная литература - 8.[4,5], 9.[24]</i></p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарским и практическим занятиям; -выполнение домашних заданий</p>
Обработка потоковых данных	<p>Spark Streaming. Преобразования без сохранения состояния. Преобразования с сохранением состояния. Операции вывода. Источники исходных данных. Основные и дополнительные источники данных.</p> <p><i>Рекомендуемые источники:</i> <i>Основная литература – 8.[2,3], 9.[24]</i> <i>Дополнительная литература - 8.[4,5], 9.[24]</i></p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарским и практическим занятиям; -выполнение домашних заданий</p>

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Таблица 5

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Подходы к обработке больших данных	Сборка автономных приложений. Управление распределением данных. Определение объекта управления распределением. Работа с файловыми системами: локальная файловая система, HDFS.	Работа с текстом лекции, разбор вопросов и заданий по теме занятия; изучение рекомендованных к занятию литературных источников.
Обработка данных в Apache Spark	Настройка отказоустойчивости. Копирование в контрольных точках. Повышение отказоустойчивости драйвера. Повышение отказоустойчивости рабочих узлов Spark Streaming.	Работа с текстом лекции, разбор вопросов и заданий по теме занятия; изучение рекомендованных к занятию литературных источников.
SparkSQL	Широковещательные переменные. Работа с разделами по отдельности. Настройка Spark с помощью SparkConf. Производительность SparkSQL.	Работа с текстом лекции, разбор вопросов и заданий по теме занятия; изучение рекомендованных к занятию литературных источников.
Обработка потоковых данных	Настройка отказоустойчивости. Копирование в контрольных точках. Повышение отказоустойчивости драйвера. Повышение отказоустойчивости рабочих узлов Spark Streaming.	Работа с текстом лекции, разбор вопросов и заданий по теме занятия; изучение рекомендованных к занятию литературных источников.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный вариант контрольной работы

Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Каждое приложение характеризуется наименованием, категорией, целевой аудиторией, рейтингом, количеством загрузок и т.д. Необходимо средствами Spark загрузить и обработать набор данных:

1. Вывести название категории с наименьшим количеством приложений.
2. Вывести название категории, минимальный рейтинг приложения, максимальный рейтинг приложения для этой категории.
3. Вывести список (Название категории, Среднее значение количества отзывов, Среднее значение количества скачиваний).
4. Вывести 5 приложений для каждой категории с наибольшим рейтингом, с указанием: Категории, Названия приложения, Рейтинга, Количества скачиваний) Отсортировать по рейтингу.
5. Вывести по два приложения для каждой категории: с максимальным количеством скачиваний, с максимальным рейтингом. В выходных данных указать: Название категории, Название приложения, Рейтинг, Количество скачиваний).

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в п.2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Таблица 6

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания

НП «Обработка больших данных и разработка интеллектуальных приложений»

Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (ОПК-1)	1. Обладает широким кругозором в области информатики, экономики, математики, что позволяет самостоятельно формулировать и решать прикладные профессиональные задачи в сфере информационных технологий	<p>Знать: принципы объединения данных из различных источников информации для последующего их анализа и обработки.</p> <p>Уметь: формулировать и решать задачи обработки больших данных; уметь обращаться к внешним ресурсам для получения данных.</p>	<p>Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Вывести названия приложений со средним значением взвешенного показателя (Sentiment_Polarity* Sentiment_Subjectivity) превышающим 0.5.</p> <p>Загрузить данные о финансовой активности клиентов банка из файлов в форматах JSON, CSV, TXT в Spark RDD для последующей обработки.</p>
	2.Использует современные методы самообучения для получения необходимых профессиональных знаний, в том числе реализованные с помощью современных технологий для решения поставленных нестандартных задач.	<p>Знать: основные принципы обработки статистических потоковых больших данных.</p> <p>Уметь: использовать специализированное программное обеспечение для решения задач распределенной обработки больших объемов информации.</p>	<p>Сформулировать структуру распределенной системы обработки больших данных по заданным критериям.</p> <p>Для набора данных [“This is an error message”, “This is an info message”, “This is a success message”, “Consider this a warning”] средствами Spark вывести количество элементов, содержащих информацию об ошибке или предупреждение.</p>
	3.Применяет знания, получаемые в ходе самообучения, для формализации новых зависимостей	<p>Знать: области применения технологий обработки больших данных, принципы</p>	<p>Доложите подходы и принципы построения отказоустойчивого кластера Hadoop.</p>

	<p>между изучаемыми явлениями и процессами и для достижения практических результатов в новой или незнакомой среде.</p>	<p>построения систем массово-параллельной обработки данных.</p> <p>Уметь: проектировать структуры для обработки больших объемов информации при решении специализированных задач в различных прикладных областях.</p>	<p>Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Создать средствами SparkSQL ETL-pipeline для последующей статистической обработки набора данных.</p>
<p>Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач (ОПК-2)</p>	<p>1.Обладает практическими навыками по созданию и разработке алгоритмов и программных средств для решения профессиональных задач.</p>	<p>Знать: принципы адаптации и сопряжения алгоритмов анализа данных и архитектуры хранилища данных.</p> <p>Уметь: применять практические навыки по созданию и разработке алгоритмов обработки потоковых больших данных.</p>	<p>Назовите источники данных приложения Spark для пакетной и интерактивной обработки</p> <p>Задан набор данных [1,2,1,3,4,1,1]. Получить средствами Spark.aggregate кортеж из двух значений вида: (произведение всех элементов списка, максимальное значение элементов списка).</p>
	<p>2.Достигает конкретных результатов на основе алгоритмизации и автоматизации решения профессиональных задач с использованием современных информационных технологий.</p>	<p>Знать: основные алгоритмы и методики обработки статистических потоковых больших данных.</p> <p>Уметь: осуществлять импорт данных из различных баз данных и хранилищ данных для их последующей</p>	<p>Сформулировать структуру приложения, которое в потоковом режиме производит анализ журнала обращения к серверу и агрегирует информацию об ошибках.</p> <p>Загрузить данные в соответствии с заданной предметной областью в Spark из базы данных HBase</p>

		автоматизированной обработки.	
	3. Демонстрирует глубокие знания по широкому спектру современных информационных технологий, необходимых для алгоритмизации и программирования прикладных задач.	<p>Знать: инструментальные и программные средства аналитической обработки данных.</p> <p>Уметь: использовать программные пакеты обработки больших данных для решения прикладных задач профессиональной деятельности.</p>	<p>Какие программы входят в состав экосистемы Hadoop, их функциональное назначение.</p> <p>Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Вывести в процентном соотношении количество приложений, имеющих средний взвешенный показатель: до 0.25, до 0.5, до 0.75, выше 0.75</p>
НП: «Управление большими данными»			
Способность анализировать предметную область и исследовать ИТ-рынок для решения задач информатизации, готовить аналитические обзоры с обоснованными выводами и рекомендациями (ПКН-6)	1. Анализирует предметную область и ИТ-рынок для решения задач информатизации.	<p>Знать: принципы объединения данных из различных источников информации для последующего их анализа и обработки.</p> <p>Уметь: формулировать и решать задачи обработки больших данных; уметь обращаться к внешним ресурсам для получения данных.</p>	<p>Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Вывести названия приложений со средним значением взвешенного показателя (Sentiment_Polarity* Sentiment_Subjectivity) превышающим 0.5.</p> <p>Загрузить данные о финансовой активности клиентов банка из файлов в форматах JSON, CSV, TXT в Spark RDD для последующей обработки.</p>

	<p>2.Готовит аналитические обзоры обоснованными выводами рекомендациями.</p>	<p>Знать: основные принципы обработки статистических потоковых больших данных.</p> <p>Уметь: готовить аналитические обзоры с обоснованными выводами использования специализированное программное обеспечение для решения задач распределенной обработки больших объемов информации.</p>	<p>Для набора данных [“This is an error message”, “This is an info message”, “This is a success message”, “Consider this a warning”] средствами Spark вывести количество элементов, содержащих информацию об ошибке или предупреждение.</p> <p>Набор данных «Google App Store» содержит информацию о приложениях, доступных для скачивания на мобильные устройства. Создать средствами SparkSQL ETL-pipeline для последующей статистической обработки набора данных.</p>
--	--	---	--

Примерные вопросы для подготовки к зачету / экзамену

1. Понятие больших данных. Характеристики больших данных
2. Экосистема Hadoop. Принципы построения отличительные особенности
3. Парадигма MapReduce.
4. Общая структура кластера Hadoop.
5. Принципы распределенного хранения данных в Hadoop.
6. Принципы организации распределенных вычислений в Hadoop.
7. Компоненты экосистемы Hadoop.
8. Файловая система HDFS. Общие понятия, особенности.
9. YARN. Отличительные особенности.
10. Структура кластера YARN. Задачи ResourceManager.
11. Структура кластера YARN. Задачи ApplicationMaster.
12. Этапы выполнения приложений в кластере YARN.

13. Microsoft Azure. Сервис HDInsight.
14. Фреймворк Spark. Особенности реализации.
15. Структурные компоненты Spark.
16. RDD. Абстракции Spark.
17. Преобразования и действия над RDD Spark.
18. Архитектуры и абстракции потоковой обработки больших данных.
19. Spark Streaming. Преобразования без сохранения состояния.
20. Spark Streaming. Преобразования с сохранением состояния.
21. Spark Streaming. Операции вывода. Источники исходных данных.
22. Основные и дополнительные источники данных.
23. Отказоустойчивость рабочих узлов и приемников.
24. Проблемы производительности.
25. Интервал пакетирования и протяженность окна.
26. Степень параллелизма.
27. Сборка мусора и использование памяти.

Пример экзаменационного билета

Федеральное государственное образовательное бюджетное учреждение
высшего образования

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Департамент анализа данных и машинного обучения

Дисциплина **Обработка статистических потоковых больших данных**

Факультет **Информационных технологий и анализа больших данных**

Форма обучения **заочная**

Направление подготовки **09.04.03 - Прикладная информатика**

Направленность программы: _____

Учебный ___/___ год

_____ модуль

Экзаменационный билет №

1. Архитектуры и абстракции потоковой обработки больших данных. (15 баллов) (ОПК-1)
2. Экосистема Hadoop. Принципы построения отличительные особенности. (15 баллов) (ОПК-2)
3. Задача. Для набора данных [“This is an error message”, “This is an info message”, “This is a success message”, “Consider this a warning”] вывести количество элементов, содержащих

информацию об ошибке или предупреждение. (30 баллов) (ОПК-1, ОПК-2)

Подготовил:



М.С. Чипчагов

Первый заместитель руководителя департамента:

В.Г. Феклин

_____Дата

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Миркин, Б. Г. Введение в анализ данных: учебник и практикум для бакалавриата и магистратуры / Б. Г. Миркин. – Москва : Юрайт, 2019. - 174 с. – Текст : непосредственный. – То же. - 2022. — ЭБС Юрайт. — URL: <https://urait.ru/bcode/469306> (дата обращения: 14.10.2022). — Текст : электронный.

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Юрайт, 2022. — 490 с. — (Высшее образование). — ЭБС Юрайт. — URL: <https://urait.ru/bcode/489100> (дата обращения: 14.10.2022). — Текст: электронный.

3. Мартишин, С. А. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala: учебное пособие / С. А. Мартишин, В. Л. Симонов, М. В. Храпченко. — Москва : ИНФРА-М, 2021. — 235 с. — ЭБС ZNANIUM.com. — URL: <http://znanium.com/catalog/product/1189321> (дата обращения: 14.10.2022). - Текст : электронный.

Дополнительная литература:

4. Точилкина, Т. Е. Хранилища данных и средства бизнес-аналитики = Data warehouse and business analytics tools: учебное пособие / Т. Е. Точилкина, А. А. Громова; Финуниверситет, Каф. "Бизнес - информатика". - Москва: Финуниверситет, 2017. - 161 с. - 1 CD. - ЭБ Финуниверситета. - URL:

http://elib.fa.ru/fbook/tochilkina_1827.pdf (дата обращения: 14.10.2022). – Текст : электронный.

5. Потемкин, А. В. Анализ данных: учебное пособие = Data analysis. Tutorial / А. В. Потемкин, И. М. Эйсымонт. – Москва : Финуниверситет, 2014. - 159 с. – Текст : непосредственный. — То же. - ЭБ Финуниверситета. — URL: http://elib.fa.ru/rbook/potemkin_asymont.pdf (дата обращения: 14.10.2022). - Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Личный кабинет обучающегося <https://org.fa.ru>
2. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
3. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
4. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
5. Электронно-библиотечная система Znanium <http://www.znaniy.com>
6. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
7. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
8. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
9. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. Национальная электронная библиотека <http://нэб.рф/>
12. Ресурсы информационно-аналитического агентства по финансовым рынкам Cbonds.ru <https://cbonds.ru/>
13. СПАРК <https://spark-interfax.ru/>
14. Academic Reference <http://ar.cnki.net/ACADREF>
15. Bank Focus <http://library.fa.ru/resource.asp?id=527>
16. Пакет баз данных компании EBSCO Publishing, крупнейшего агрегатора научных ресурсов ведущих издательств мира <http://search.ebscohost.com>

17. Emerald: Management eJournal Portfolio <https://www.emerald.com/insight/>
18. Информационно-аналитическая база данных EMIS Global <https://www.emis.com/php/companies/overview/index>
19. Реферативная база данных по математике MathSciNET <https://mathscinet.ams.org/mathscinet/>
20. Oxford Scholarship Online <https://oxford.universitypressscholarship.com/>
21. Коллекция научных журналов Oxford University Press <https://academic.oup.com/journals/>
22. Scopus <https://www.scopus.com>
23. Электронная коллекция книг издательства Springer: Springer eBooks <http://link.springer.com/>
24. Документация Spark <http://spark.apache.org/docs/latest/>

10. Методические указания для обучающихся по освоению дисциплины

Самостоятельная работа студентов проходит внеаудиторно. Организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения.

Домашние задания следует выполнять регулярно при подготовке к практическим занятиям. Контроль выполнения домашних заданий осуществляется в ходе практических занятий в процессе выборочного собеседования.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

Пакет офисных программ.

Антивирус Kaspersky.

11.2. Современные профессиональные базы данных и информационные справочные системы:

Информационно-правовая система «Консультант Плюс»;

Информационно-правовая система «Гарант»;

Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

Система комплексного раскрытия информации «СКРИН»: <https://skrin.ru>

11.3. Сертифицированные программные и аппаратные средства защиты информации не предусмотрены.

11.4. Apache Spark.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для освоения дисциплины необходимо использование учебных аудиторий, оборудованных вычислительной техникой с доступом к сети Интернет.