

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

**Кафедра информационных технологий
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

Проректор по учебной
и методической работе

_____ Е.А. Каменева

28.05.2025 г.

Абаев В.А.

ОБРАБОТКА СТАТИЧЕСКИХ И ПОТОКОВЫХ БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

38.04.02 - Менеджмент,

Направленность программы:

"Логистика: финансовые и цифровые технологии"

Рекомендовано Ученым советом

*Факультета информационных технологий и анализа больших данных
(протокол № 55 от 20 мая 2025 г.)*

*Одобрено советом Кафедры информационных технологий
(протокол № 4 от 19 мая 2025 г.)*

Москва 2025

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.....	3
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	4
5.1. Содержание дисциплины.....	4
5.2. Учебно–тематический план.....	6
5.3. Содержание семинаров, практических занятий.....	6
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	8
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	8
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	9
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	12
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	15
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	15
10. Методические указания для обучающихся по освоению дисциплины	17
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	17
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	18

1. Наименование дисциплины

«Обработка статических и потоковых больших данных».

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

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПК-6	Способность анализировать предметную область и исследовать ИТ-рынок для решения задач информатизации, готовить аналитические обзоры с обоснованными выводами и рекомендациями	1. Владеет навыками для анализа и отбора информационных технологий при решении профессиональных задач.	Знать: методы и технологии анализа и отбора статистических и потоковых больших данных, развертывания информационной системы обработки статистических и потоковых больших данных Уметь: реализовывать решения формирования, обработки и анализа статистических и потоковых больших данных
		2. Демонстрирует умение создавать аналитические обзоры с обоснованными выводами и рекомендациями	Знать: методы и шаблоны решений по формированию аналитических обзоров, отчетов, аналитических панелей по результатам обработки статистических и потоковых больших данных Уметь: обобщать и визуализировать результаты обработки статистических и потоковых больших данных, оценивать производительность информационной системы, давать рекомендации по ее модификации.

3. Место дисциплины в структуре образовательной программы

Дисциплина «Обработка статических и потоковых больших данных» относится к Модулю дисциплин дополнительной квалификации – специалист по автоматизации бизнес-процессов по направлению подготовки 38.04.01 – Менеджмент, направленность программы: «Логистика: финансовые и цифровые технологии».

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

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Модуль 6 (в часах)
Общая трудоёмкость дисциплины	3/108	108
<i>Контактная работа - Аудиторные занятия</i>	<i>30</i>	<i>30</i>
<i>Лекции</i>	<i>10</i>	<i>10</i>
<i>Семинары, практические занятия</i>	<i>20</i>	<i>20</i>
<i>Самостоятельная работа</i>	<i>78</i>	<i>78</i>
Вид текущего контроля		Домашнее творческое задание
Вид промежуточной аттестации		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. Pyspark MLlib

Классификация с использованием Pyspark ML. ВекторАссемблер, оценщики и преобразователи. RФормула, импутация PySpark. Кроссвалидатор PySpark. Дерево решений PySpark, случайный лес PySpark. Оценщик PySpark, гиперопт с Pyspark.

Тема 5. Pyspark GraphFrames

Фрейм данных вершин и фрейм данных ребер. Постройте неориентированного графа. Построение сюжетно-ориентированного графа. Определение степени узлов графа. Поиск простых MOTIFS в GraphFrames. Поиск в ширину. Функция кратчайших путей графа Pyspark GraphFrames.

Тема 6. 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

№ п/ п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа- Аудиторная работа			Самост оательн ая работа	
			Об- щая, в т.ч.:	Лек ции	Семинары, практические занятия		
1.	Подходы к обработке больших данных	15	3	1	2	12	Самостоятельные работы. Участие в решении задач на практических занятиях. Обсуждение решенных задач.
2.	Обработка данных в Apache Spark	15	3	1	2	12	
3.	SparkSQL	18	6	2	4	12	
4.	Pyspark MLlib	20	6	2	4	14	
5.	Pyspark GraphFrames	20	6	2	4	14	
6.	Обработка поточковых данных	20	6	2	4	14	
	В целом по дисциплине	108	30	10	20	78	Согласно учебному плану: домашнее творческое задание
	Итого в %		28	33	67	72	

* объем контактной работы в очно-заочной/заочной формах обучения и индивидуальных учебных планах определяется соответствующими учебными планами. Темы, реализуемые в виде контактной работы, определяются преподавателем самостоятельно, исходя из уровня их сложности.

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9	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. Форматы Hadoop для ввода и вывода. Сжатие файлов. Рекомендуемые источники: 8.[1,2,4,5], 9.[24]	-работа с текстом лекции; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарским и практическим занятиям; - выполнение домашних заданий

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>Рекомендуемые источники: 8.[2,3,5], 9.[24]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарским и практическим занятиям;</p> <p>-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>Рекомендуемые источники: 8.[1,2,3,4,5], 9.[24]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарским и практическим занятиям;</p> <p>-выполнение домашних заданий</p>
Pyspark MLlib	<p>Классификация с использованием Pyspark ML. ВекторАссемблер, оценщики и преобразователи. RФормула, импутация PySpark. Кроссвалидатор PySpark. Дерево решений PySpark, случайный лес PySpark. Оценщик PySpark, гиперопт с Pyspark.</p> <p>Рекомендуемые источники: 8.[1,2,3,4,5], 9.[24]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарским и практическим занятиям;</p> <p>-выполнение домашних заданий</p>
Pyspark GraphFrames	<p>Фрейм данных вершин и фрейм данных ребер. Постройте неориентированного графа Построение сюжетно-ориентированного графа. Определение степени узлов графа. Поиск простых MOTIFS в GraphFrames. Поиск в ширину. Функция кратчайших путей графа Pyspark GraphFrames.</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию</p>

	Рекомендуемые источники: 8.[1,2,3,4,5], 9.[24]	литературных источников; -подготовка к семинарским и практическим занятиям; -выполнение домашних заданий
Обработка потоковых данных	Spark Streaming. Преобразования без сохранения состояния. Преобразования с сохранением состояния. Операции вывода. Источники исходных данных. Основные и дополнительные источники данных. Рекомендуемые источники: 8.[2,3,4,5], 9.[24]	-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарским и практическим занятиям; -выполнение домашних заданий

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

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

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.

Pyspark MLlib	Корреляция, Проверка гипотезы, ChiSquareTest, Сумматор, Источник данных LIBSVM, Источник данных изображения, Конвейеры машинного обучения	Работа с текстом лекции, разбор вопросов и заданий по теме занятия; изучение рекомендованных к занятию литературных источников.
Pyspark GraphFrames	Теория и концепции графов. Графовые платформы и обработка. Алгоритмы поиска по графу и поиска пути. Алгоритмы вычисления центральности. Алгоритмы выделения сообществ. Графовые алгоритмы на практике. Графовые алгоритмы и машинное обучение	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 домашнего творческого задания

Проэксплуатировать, используя свои наборы данных, результаты представить в виде блокнота Colab или Jupyter сопроводив комментариями и результатами:

1.Объединение двух RDD

```

from pyspark import SparkContext, SparkConf
appName = 'appName'
master = 'local[*]'
sc.stop()
conf = SparkConf().setAppName(appName).setMaster(master)
sc = SparkContext(conf=conf)
R = sc.textFile("d:/R.txt");
print(R.collect())
S = sc.textFile("d:/S.txt");
print(S.collect())
print(type(R))
print(type(S))
r1 = R.map(lambda s: s.split(", "))
print(r1.collect())

```

```

r2 = r1.flatMap(lambda s: [(s[0], s[1])])
print(r2.collect())
s1 = S.map(lambda s: s.split(", "))
print(s1.collect())
s2 = s1.flatMap(lambda s: [(s[0], s[1])])
print(s2.collect())
RjoinedS = r2.join(s2)
RjoinedS.collect()

```

```

['k1,v1', 'k1,v2', 'k2,v3', 'k2,v4', 'k3,v7', 'k3,v8', 'k3,v9']
<class 'pyspark.rdd.RDD'>
<class 'pyspark.rdd.RDD'>
['k1,v11', 'k1,v22', 'k1,v33', 'k2,v55', 'k4,v77', 'k5,v88']
[['k1', 'v1'], ['k1', 'v2'], ['k2', 'v3'], ['k2', 'v4'], ['k3', 'v7'], ['k3', 'v8'], ['k3', 'v9']]
[('k1', 'v1'), ('k1', 'v2'), ('k2', 'v3'), ('k2', 'v4'), ('k3', 'v7'), ('k3', 'v8'), ('k3', 'v9')]
[['k1', 'v11'], ['k1', 'v22'], ['k1', 'v33'], ['k2', 'v55'], ['k4', 'v77'], ['k5', 'v88']]
[('k1', 'v11'), ('k1', 'v22'), ('k1', 'v33'), ('k2', 'v55'), ('k4', 'v77'), ('k5', 'v88')]
[('k1', ('v1', 'v11')),
 ('k1', ('v1', 'v22')),
 ('k1', ('v1', 'v33')),
 ('k1', ('v2', 'v11')),
 ('k1', ('v2', 'v22')),
 ('k1', ('v2', 'v33')),
 ('k2', ('v3', 'v55')),
 ('k2', ('v4', 'v55'))]

```

2. Использование map() в RDD

```

from pyspark import SparkContext, SparkConf
appName = 'appName'
master = 'local[*]'
sc.stop()
conf = SparkConf().setAppName(appName).setMaster(master)
sc = SparkContext(conf=conf)
nums = sc.parallelize([1, 2, 3, 4, 5])
print(nums.collect())

```

```
bytwo = nums.map(lambda x: x + 2)
print(bytwo.collect())
squared = nums.map(lambda x: x * x)
print(squared.collect())
```

```
[1, 2, 3, 4, 5]
[3, 4, 5, 6, 7]
[1, 4, 9, 16, 25]
```

3. Использование filter() в RDD

```
from pyspark import SparkContext, SparkConf
appName = 'appName'
master = 'local[*]'
sc.stop()
conf = SparkConf().setAppName(appName).setMaster(master)
sc = SparkContext(conf=conf)
nums = sc.parallelize([1, 2, 3, 4, 5, 6, 7])
print(nums.collect())
filtered1 = nums.filter(lambda x : x % 2 == 1)
print(filtered1.collect())
filtered2 = nums.filter(lambda x : x % 2 == 0)
print(filtered2.collect())
```

```
[1, 2, 3, 4, 5, 6, 7]
[1, 3, 5, 7]
[2, 4, 6]
```

4. Среднее значение в RDD

```
from pyspark import SparkContext, SparkConf
appName = 'appName'
master = 'local[*]'
sc.stop()
conf = SparkConf().setAppName(appName).setMaster(master)
sc = SparkContext(conf=conf)
nums = sc.parallelize([1, 2, 3, 4, 5, 6, 7, 8, 20])
```

```

nums.collect()

sumAndCount = nums.map(lambda x: (x, 1)).fold((0, 0), (lambda x, y: (x[0] + y[0], x[1] + y[1])))

print(sumAndCount)

avg = float(sumAndCount[0]) / float(sumAndCount[1])

print(avg)

```

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры информационных технологий 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 разделе 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

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
Способность анализировать предметную область и исследовать ИТ-рынок для решения задач информатизации, готовить аналитические обзоры с обоснованными выводами и рекомендациями ПК-6	1. Владеет навыками для анализа и отбора информационных технологий при решении профессиональных задач.	Знать: методы и технологии анализа и отбора статистических и потоковых больших данных, развертывания информационной системы обработки статистических и потоковых больших данных	Создать PySpark DataFrame из списка строк Создать PySpark DataFrame с явной схемой Создать PySpark DataFrame из pandas DataFrame Создать PySpark DataFrame из RDD, состоящего из списка кортежей

		<p>Уметь: реализовывать решения формирования, обработки и анализа статистических и поточковых больших данных</p>	<p>Просмотр данных (вывод схемы) Просмотр данных (вывод статистики) Просмотр данных (DataFrame.take() или DataFrame.tail()) Выбор и доступ к данным (возвращает экземпляр Column) Выбор и доступ к данным (DataFrame.select()) Выбор и доступ к данным (DataFrame.filter()) Применение функции (использование API- интерфейсов в серии pandas) Создание объекта (Создание Spark DataFrame из pandas DataFrame) Создание объекта (Создание pandas-on- Spark DataFrame из Spark DataFrame)</p>
	<p>2. Демонстрирует умение создавать аналитические обзоры с обоснованными выводами и рекомендациями</p>	<p>Знать: методы и шаблоны решений по формированию аналитических обзоров, отчетов, аналитических панелей по результатам обработки статистических и поточковых больших данных Уметь: обобщать и визуализировать результаты обработки статистических и поточковых больших данных, оценивать производительность информационной системы, давать рекомендации по ее модификации.</p>	<p>Группировка данных PySpark DataFrame Создание объекта (Создание серии pandas-on-Spark) Создание объекта (Создание кадра данных pandas-on- Spark) Создание объекта (Создание pandas DataFrame путем передачи массива numpy с индексом даты и времени и помеченными столбцами) Создание объекта (кадр данных pandas преобразовать в кадр данных pandas-on- Spark) Создание объекта (Отображение индекса,</p>

			столбцов и базовых данных mapreduce) Транспонирование ваших данных Сортировка по индексу Сортировка по значению Группировка и последующее применение функции sum() к полученным группам
--	--	--	---

Примерные вопросы для подготовки к зачету

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.

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

Основная литература:

1. Миркин, Б. Г. Базовые методы анализа данных: учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва: Юрайт, 2025. — 297 с. — ЭБС Юрайт. — URL: <https://urait.ru/index.php/bcode/560414> (дата обращения: 31.03.2025). — Текст: электронный.
2. Анализ данных: учебник для вузов / под редакцией В. С. Мхитаряна. — Москва: Юрайт, 2025. — 448 с. — (Высшее образование). — ЭБС Юрайт. — URL: <https://www.urait.ru/bcode/560311> (дата обращения: 31.03.2025). — 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, 2024. — 235 с. - ЭБС ZNANIUM. - URL: <https://znanium.ru/catalog/product/2139860> (дата обращения: 31.03.2025). - 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. — Москва: Финуниверситет, 2017. - 161 с. - 1 CD. - ЭБ Финуниверситета. - URL: http://elib.fa.ru/fbook/tochilkina_1827.pdf (дата обращения: 31.03.2025). — 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: 31.03.2025). - 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.