

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ**  
Обработка больших объемов данных

**Код модуля**  
1153907

**Модуль**  
Базы данных

**Екатеринбург**

Оценочные материалы составлены автором(ами):

<b>№ п/п</b>	<b>Фамилия, имя, отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Корелин Иван Андреевич	кандидат технических наук	Доцент	Базовая кафедра "Автоматизация финансовых систем"

**Согласовано:**

Управление образовательных программ

Т.Г. Комарова

**Авторы:**

- Корелин Иван Андреевич, Доцент, Базовая кафедра "Автоматизация финансовых систем"

**1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Обработка больших объемов данных**

1.	Объем дисциплины в зачетных единицах	6	
2.	Виды аудиторных занятий	Лекции Практические/семинарские занятия	
3.	Промежуточная аттестация	Экзамен	
4.	Текущая аттестация	Контрольная работа	1
		Домашняя работа	1

**2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Обработка больших объемов данных**

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Планируемые результаты обучения (индикаторы)	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ПК-1 -Способен использовать методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации посредством информационных технологий	<p>З-1 - Изложить основные методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации</p> <p>З-2 - Характеризовать особенности применения информационных технологий в области получения, передачи, хранения, переработки и представления информации</p> <p>П-1 - Осуществлять обоснованный выбор методов цифровой обработки сигналов для переработки и представления информации посредством информационных технологий с учетом</p>	<p>Домашняя работа</p> <p>Контрольная работа</p> <p>Лекции</p> <p>Практические/семинарские занятия</p> <p>Экзамен</p>

	<p>особенностей сигналов и изображений</p> <p>У-1 - Выбирать адекватные методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации</p> <p>У-2 - Систематизировать и оценивать научно-техническую информацию о методах цифровой обработки сигналов для переработки и представления информации</p>	
<p>ПК-7 -Способен разрабатывать системы управления базами данных</p>	<p>З-2 - Классифицировать основные структуры данных</p> <p>З-3 - Характеризовать методы обработки данных</p> <p>П-2 - Сделать вывод об оптимальности применения методов обработки данных для системы управления базами данных</p> <p>У-2 - Определять оптимальные методы обработки данных</p>	<p>Домашняя работа</p> <p>Контрольная работа</p> <p>Лекции</p> <p>Практические/семинарские занятия</p> <p>Экзамен</p>

### 3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

#### 3.1. Процедуры текущей и промежуточной аттестации по дисциплине

<p><b>1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.50</b></p>		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	3,17	50
<i>контрольная работа</i>	3,10	50
<p><b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.50</b></p>		
<p><b>Промежуточная аттестация по лекциям – экзамен</b></p>		
<p><b>Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.50</b></p>		
<p><b>2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.50</b></p>		

Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение практических работ</i>	3,17	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– <b>1.00</b>		
Промежуточная аттестация по практическим/семинарским занятиям– <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– <b>0.00</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –не предусмотрено</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по лабораторным занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – <b>не предусмотрено</b>		
<b>4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено</b>		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по онлайн-занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – <b>не предусмотрено</b>		

### 3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– <b>не предусмотрено</b>		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – <b>не предусмотрено</b>		

## 4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

**Критерии оценивания учебных достижений обучающихся**

<b>Результаты обучения</b>	<b>Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам</b>
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

**Шкала оценивания достижения результатов обучения (индикаторов) по уровням**

<b>Характеристика уровней достижения результатов обучения (индикаторов)</b>				
<b>№ п/п</b>	<b>Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)</b>	<b>Шкала оценивания</b>		
		<b>Традиционная характеристика уровня</b>		<b>Качественная характеристика уровня</b>
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам,	Неудовлетворитель но	Не зачтено	Недостаточный (Н)

	имеются существенные ошибки и замечания, требуется доработка	(менее 40 баллов)	
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания	Нет результата

## 5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

### 5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

#### 5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

#### 5.1.2. Практические/семинарские занятия

Примерный перечень тем

1. Реализация приложения DataGenerator
  2. Подсчёт частоты встречаемости слов WordCount (single computer)
  3. Подсчёт слов с Hadoop MapReduce WordCount (Hadoop cluster)
  4. Расчет рейтингов фильмов
  5. Подсчёт частоты встречаемости слов с Apache Spark RDD
  6. Работа с данными в Spark DataFrameAPI
  7. Работа с источниками данных в Apache Spark Очистка и подготовка данных.
  8. Анализ данных в Apache Spark.
  9. Использование SQL в Apache Spark.
- LMS-платформа – не предусмотрена

### 5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

#### Базовый

#### 5.2.1. Контрольная работа

Примерный перечень тем

1. Работа с данными в различных форматах.

Примерные задания

Примерные задания в составе контрольных работ:

1. Формат данных CSV может быть использован как замена:
  - реляционным СУБД
  - нереляционным СУБД
  - документо-ориентированным СУБД

2. Что обычно выступает разделителем столбцов в формате CSV:

- Запятая
- Точка с запятой
- Перенос строки

3. Какие элементы не используются в формате JSON в качестве структурных элементов:

- Теги
- Кавычки
- Двоеточие

4. Для какого языка программирования впервые использовался формат JSON:

- JavaScript
- Java
- Python

5. В каком виде хранятся данные в MongoDB:

- BSON
- JSON
- XML

6. Выберите наиболее подходящие характеристики MongoDB:

- Ключ-значение, неструктурированность данных, нереляционные свойства
- Ключ-значение, неструктурированность данных, реляционные свойства
- SQL, неструктурированность данных, реляционные свойства

LMS-платформа – не предусмотрена

### 5.2.2. Домашняя работа

Примерный перечень тем

1. Домашняя работа №1: Создание собственного набора данных.
2. Домашняя работа №2: Обработка данных в Apache Spark.

Примерные задания

1. Подготовить собственный набор данных. Выберите задачу в одном из направлений создания системы искусственного интеллекту (компьютерное зрение, обработка естественного языка) и подготовьте для этой задачи набор данных для обучения с учителем. Соберите и очистите данные, проведите разметку. Готовый набор данных разместите на одной из общедоступных платформ для хранения данных по своему выбору. Подготовьте документацию к созданному набору данных.

Перечень задач, для которых рекомендуется подготовить набор данных:

Классификация объектов на изображениях.

Определение положения объектов на изображениях.

Определение положения объектов в видео.

Классификация текста на русском языке.

Определение эмоциональной окраски текста на русском языке.



## Домашняя работа №2: Обработка данных в Apache Spark.

Создайте набор данных в Apache Spark и проведите его исследование с помощью Spark DataFrame API.

Схема данных выглядит следующим образом: см. файл во вложении.

Онлайн-школа продает образовательные продукты: онлайн-курсы, книги, семинары и т.п. Описание и стоимость продуктов содержится в таблице Products. Когда клиент что-то покупает, создается заказ, который заносится в таблицу Orders. Заказ может содержать несколько продуктов, перечень продуктов в заказах содержится в таблице Sales.

Таблица Products - продукты онлайн-школы:

- id - идентификатор продукта
- name - название продукта
- price - стоимость продукта

Таблица Orders - заказы:

- id - идентификатор заказа
- order\_date - дата заказа
- customer\_id - идентификатор заказчика (таблица с заказчиками не создается для упрощения примера)

Таблица Sales - продажи:

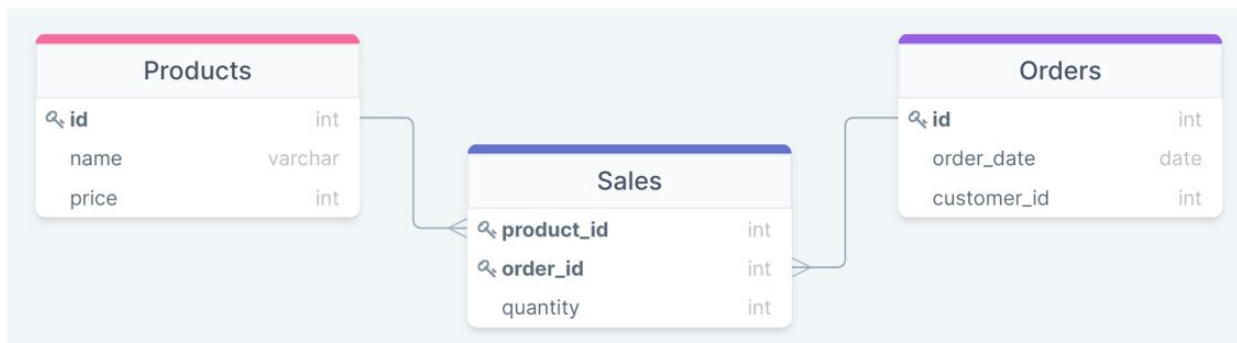
- product\_id - идентификатор продукта, ссылка на таблицу Products, поле id
- order\_id - идентификатор заказа, ссылка на таблицу Orders, поле id
- quantity - количество продуктов в заказе

Ноутбук в облачной платформе Colab с заготовкой кода для домашней работы – [https://colab.research.google.com/drive/1MLiIHIZ2CcBbCp\\_U7wKs\\_CAPmnx2F7Q6?usp=sharing](https://colab.research.google.com/drive/1MLiIHIZ2CcBbCp_U7wKs_CAPmnx2F7Q6?usp=sharing)

Задания для анализа:

- Выведите список продуктов, которые не были проданы ни разу
- Определите, сколько продуктов любого типа было продано по дням.
- Определить, какая выручка от продуктов любого типа была получена по дням.

Схема данных выглядит следующим образом:



LMS-платформа – не предусмотрена

### 5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

#### 5.3.1. Экзамен

Список примерных вопросов

1. Архитектура центров обработки данных.
2. Кластеры для параллельных и распределенных вычислений.
3. Экосистема для распределенного хранения и обработки больших объемов данных:

Apache Hadoop.

4. Распределенная файловая система HDFS.
5. Общедоступные платформы для хранения данных.
6. Работа с данными в формате CSV в PySpark.
7. Работа с данными в формате JSON в PySpark.
8. Работа с данными в формате HTML в PySpark.
9. Работа с Parquet в PySpark.
10. Работа с графами знаний в Python.
11. Этапы и инструменты создания наборов данных для машинного обучения.
12. Методы и инструменты подготовки данных.
13. Методы и инструменты очистки данных.
14. Распределенная обработка данных в Apache Spark.
15. Работа с данными с использованием Apache Spark DataFrame.
16. Источники данных для Apache Spark DataFrame.
17. Обработка данных в Apache Spark DataFrame.
18. Использование SQL в Apache Spark DataFrame.

LMS-платформа – не предусмотрена

### 5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.