

**Министерство науки и высшего образования РФ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Национальный исследовательский университет «МЭИ»**

**Направление подготовки/специальность: 09.04.01 Информатика и вычислительная техника**

**Наименование образовательной программы: Цифровые технологии**

**Уровень образования: высшее образование - магистратура**

**Форма обучения: Очная**

**Оценочные материалы  
по дисциплине  
Технология обработки больших данных**

**Москва  
2022**

## ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Преподаватель

(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Зейн А.Н.
	Идентификатор	R54353a8f-ZeynAIN-7d1f3849

(подпись)

А.Н. Зейн

(расшифровка  
подписи)

## СОГЛАСОВАНО:

Руководитель  
образовательной  
программы

(должность, ученая степень, ученое  
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Вишняков С.В.
	Идентификатор	R35b26072-VishniakovSV-02810d9

(подпись)

С.В.

Вишняков

(расшифровка  
подписи)

Заведующий  
выпускающей кафедры

(должность, ученая степень, ученое  
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Вишняков С.В.
	Идентификатор	R35b26072-VishniakovSV-02810d9

(подпись)

С.В.

Вишняков

(расшифровка  
подписи)

## ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки: достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ПК-3 Способен осуществлять проектирование вычислительных комплексов и систем, включая разработку аппаратного, программного обеспечения, системную интеграцию, ввод в эксплуатацию

ИД-2 Демонстрирует знание теории баз данных, включая перспективные технологии обработки больших данных

и включает:

**для текущего контроля успеваемости:**

Форма реализации: Выполнение задания

1. Ограничения реляционной модели данных (Лабораторная работа)

Форма реализации: Защита задания

1. Лабораторная работа 1 (MongoDB часть 1) - установка MongoDB, создание документов и коллекций. (Лабораторная работа)

2. Лабораторные работы 2 и 3 - Сложные запросы для реализации Map() и Reduce а также MapReduce(). Введение в работе с аналитической системы Ktime. (Лабораторная работа)

Форма реализации: Проверка задания

1. Лабораторная работа 4 - Кластеризация и классификация в Ktime (Лабораторная работа)

## БРС дисциплины

2 семестр

Раздел дисциплины	Веса контрольных мероприятий, %				
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
	Срок КМ:	4	8	12	16
Данные: статические и динамические; Структурированные и мало-структурированные данные. Недостатки реляционных баз данных.					
Данные: статические и динамические; Структурированные и мало-структурированные данные.	+				
Недостатки реляционных баз данных	+				
Источники больших данных. DOM-модель интернет страницы.					
Источники больших данных.	+				

DOM-модель интернет страницы	+			
Копирование больших объемов данных. XML и JSON.				
Копирование больших объемов данных.		+		
XML и JSON		+		
Критерий Тьюки. Принцип Бонферрони. MapReduce.				
Критерий Тьюки. Принцип Бонферрони.		+		
MapReduce.		+		
Работа с базой данных MongoDB.				
MongoDB. Установка и работа в диалоговом режиме и в интерактивном режиме.			+	
Перенос данных из MongoDB в реляционную базу данных.				
Перенос данных между MongoDB и реляционной базы данных.			+	
Алгоритмы классификации				
Алгоритмы классификации				+
Алгоритмы кластеризации				
Алгоритмы кластеризации				+
Некоторые типовые и перспективные задачи обработки больших данных.				
Некоторые типовые и перспективные задачи обработки больших данных				+
Вес КМ:	20	20	30	30

\$Общая часть/Для промежуточной аттестации\$

## СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

### *I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций*

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ПК-3	ИД-2ПК-3 Демонстрирует знание теории баз данных, включая перспективные технологии обработки больших данных	<p>Знать:</p> <p>основные методы, обработки, анализа и интерпретации данных в системах Big Data и принцип миграции данных в гетерогенных системах</p> <p>ограничения реляционной модели при работе с большими данными</p> <p>Уметь:</p> <p>использовать существующие алгоритмы для классификации объектов и раскрывать не очевидные закономерности</p> <p>получать большие данные из открытых и доступных источников, их обрабатывать и модифицировать для работы в реляционной/не реляционной БД</p>	<p>Ограничения реляционной модели данных (Лабораторная работа)</p> <p>Лабораторная работа 1 (MongoDB часть1) - установка MongoDB, создание документов и коллекций. (Лабораторная работа)</p> <p>Лабораторные работы 2 и 3 - Сложные запросы для реализации Map() и Reduce а также MapReduce(). Введение в работе с аналитической системы Ktime. (Лабораторная работа)</p> <p>Лабораторная работа 4 - Кластеризация и классификация в Ktime (Лабораторная работа)</p>

## II. Содержание оценочных средств. Шкала и критерии оценивания

### КМ-1. Ограничения реляционной модели данных

**Формы реализации:** Выполнение задания

**Тип контрольного мероприятия:** Лабораторная работа

**Вес контрольного мероприятия в БРС:** 20

**Процедура проведения контрольного мероприятия:** выполняется лабораторная работа в рамках, которой сравниваются реляционные и нереляционные структуры данных. проводятся эксперименты. делаются выводы. Студент представляет отчет. получает вопросы.

**Краткое содержание задания:**

выполняется лабораторная работа в рамках, которой сравниваются реляционные и нереляционные структуры данных. проводятся эксперименты. делаются выводы. Студент представляет отчет. получает вопросы.

**Контрольные вопросы/задания:**

Знать: ограничения реляционной модели при работе с большими данными	1. Основные преимущества и недостатки реляционной модели данных. Основные преимущества и недостатки нереляционной модели данных. DOM модель интернет страницы. JSON - формат файла
---	---

**Описание шкалы оценивания:**

*Оценка: 5*

*Нижний порог выполнения задания в процентах: 90*

*Описание характеристики выполнения знания:* Студент отвечает на 2 вопроса с высокой точностью.

*Оценка: 4*

*Нижний порог выполнения задания в процентах: 80*

*Описание характеристики выполнения знания:* Студент отвечает на 2 вопроса. один из ответов частично полный.

*Оценка: 3*

*Нижний порог выполнения задания в процентах: 50*

*Описание характеристики выполнения знания:* Студент отвечает на 1 вопрос полноценно или на 2 вопроса частично

**КМ-2. Лабораторная работа 1 (MongoDB часть1) - установка MongoDB, создание документов и коллекций.**

**Формы реализации:** Защита задания

**Тип контрольного мероприятия:** Лабораторная работа

**Вес контрольного мероприятия в БРС:** 20

**Процедура проведения контрольного мероприятия:** выполняется лабораторная работа в рамках, которой создаётся база данные в MongoDB, создаются документы и коллекции. Студент представляет отчет. получает вопросы.

**Краткое содержание задания:**

Установить СУБД MongoDB. Создать документы и коллекции. Заполнение коллекций: добавить 1 документ, добавить несколько документов одной командой. Удалите один документ. Удалите несколько документов одной командой.

**Контрольные вопросы/задания:**

Знать: основные методы, обработки, анализа и интерпретации данных в системах Big Data и принцип миграции данных в гетерогенных системах	1. В каком виде хранятся коллекции в MongoDB? Чем отличается коллекция от документа в MongoDB? MapReduce()
---	--

**Описание шкалы оценивания:**

*Оценка: 5*

*Нижний порог выполнения задания в процентах: 80*

*Описание характеристики выполнения знания: Студент отвечает на все вопросы. один из ответов не полный.*

*Оценка: 4*

*Нижний порог выполнения задания в процентах: 70*

*Описание характеристики выполнения знания: студент отвечает на 3 из 4-х вопросов*

*Оценка: 3*

*Нижний порог выполнения задания в процентах: 50*

*Описание характеристики выполнения знания: студент отвечает на 2 из 4-х вопросов*

**КМ-3. Лабораторные работы 2 и 3 - Сложные запросы для реализации Map() и Reduce а также MapReduce(). Введение в работе с аналитической системы Kntime.**

**Формы реализации:** Защита задания

**Тип контрольного мероприятия:** Лабораторная работа

**Вес контрольного мероприятия в БРС:** 30

**Процедура проведения контрольного мероприятия:** выполняется лабораторная работа в рамках, которой работа проводится в среде MongoDB, распределяются права доступа, проводится резервное копирование и восстановление данных. Студент представляет отчет. получает вопросы.

**Краткое содержание задания:**

Написать команду для выдачи прав на чтения и запись в коллекцию Books в среде MongoDB. Реализовать функции Map(), Reduce и MapReduce. Построить простейшую диаграмму потока данных DFD в аналитической системе Kntime.

**Контрольные вопросы/задания:**

Уметь: использовать существующие алгоритмы для классификации объектов и раскрывать не очевидные закономерности	1. Написать команду для реализации функции Map(). Написать команду для реализации функции Reduce(). Написать команду для реализации функции MapReduce().
--	--

**Описание шкалы оценивания:**

*Оценка: 5*

*Нижний порог выполнения задания в процентах: 100*

*Описание характеристики выполнения знания: Студент отвечает на 2 вопроса.*

Оценка: 4

Нижний порог выполнения задания в процентах: 80

Описание характеристики выполнения знания: Студент отвечает на 2 вопроса. Один из ответов не полноценный.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Студент отвечает на 1 из 2-х вопросов.

#### **КМ-4. Лабораторная работа 4 - Кластеризация и классификация в Knime**

**Формы реализации:** Проверка задания

**Тип контрольного мероприятия:** Лабораторная работа

**Вес контрольного мероприятия в БРС:** 30

**Процедура проведения контрольного мероприятия:** Студент выполняет пункты задания. работа реализуется в аналитической системе Knime.

#### **Краткое содержание задания:**

Необходимо подготовить диаграмму потока данных и выбрать нужные компоненты для реализации кластерного анализа и классификации.

#### **Контрольные вопросы/задания:**

Уметь: получать большие данные из открытых и доступных источников, их обрабатывать и модифицировать для работы в реляционной/не реляционной БД	1. как настроить компоненты для кластерного анализа? что может быть представлено в виде входного потока данных и что в качестве выходного потока? как провести нормирование входных данных? в какой цепочке и какие компоненты нужно использовать?
--	---

#### **Описание шкалы оценивания:**

Оценка: 5

Нижний порог выполнения задания в процентах: 80

Описание характеристики выполнения знания: Студент отвечает на 4 вопроса. Один из ответов может быть не полноценный.

Оценка: 4

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Студент отвечает на 3 вопроса.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Студент отвечает на 2 вопроса полноценно.

# СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

## 2 семестр

**Форма промежуточной аттестации:** Экзамен

### Пример билета

1. Основные отличия реляционной и не реляционной БД. Когда их применить?
2. Написать команду для автоматического создания коллекции и ее заполнения одним документом.
3. Построить диаграмму DF с минимальным количеством компонентов для реализации классификации методов случайного леса в аналитической системе Knime

### Процедура проведения

Студент получает свой экзаменационный билет и готовится отвечать письменно. Допускается использование собственного ноутбука для построения диаграмм и написания команд.

### *1. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины*

**1. Компетенция/Индикатор:** ИД-2<sub>ПК-3</sub> Демонстрирует знание теории баз данных, включая перспективные технологии обработки больших данных

### Вопросы, задания

1. Когда нужно применять реляционную БД?
2. Когда нужно применять не реляционную БД?
3. Предложите несколько источников “больших данных”
4. По какой причине формат JSON используется для работы в MongoDB?
5. Что такое MapReduce() функция?
6. Какие работы нужно провести чтобы подготовить данные к анализу в Knime?
7. Что делать если наш набор данных содержит как численные так и строковые типы данных?
8. что такое A/B тестирование?
9. Какие преимущества и недостатки для метода кластеризации K-средних?
10. Какие преимущества и недостатки для метода кластеризации DBScan?

### Материалы для проверки остаточных знаний

1. За счет чего не реляционная СУБД работает “быстрее” реляционной?

Ответы:

1. 1) за счет отсутствие связей между таблицами
2. 2) за счет отсутствие ограничений
3. 3) за счет применения JSON

Верный ответ: 1 и 2

2. Почему не реляционные СУБД характеризуются как 3V системы?

Ответы:

1. 1) Volume

2. 2) Velocity
3. 3) Variety
4. 4) View
5. 5) Vector

Верный ответ: 1, 2 и 3

3. Классификация - это ....

Ответы:

1. 1) тоже самое что кластеризация
2. 2) только с применением нейронных сетей можно реализовать классификацию.
3. 3) один из методов распределения объектов по выделенным группам

Верный ответ: 3

4. Кластеризация - это ...

Ответы:

1. 1) тоже самое что классификация
2. 2) только с применением нейронных сетей можно реализовать кластеризацию
3. 3) один из методов формирования кластеров, когда нет первоначальной группировки объектов.

Верный ответ: 3

5. Функция MapReduce() состоит из 2 этапов

Ответы:

- 1) сначала Map() затем Reduce()
- 2) сначала Reduce() затем Map()
- 3) Map() и Reduce() в любом порядке. Порядок не имеет значение

Верный ответ: 1

6. Лемматизация и нормализация данных. это...

Ответы:

1. 1) для анализа данных синонимы
2. 2) разные этапы подготовки данных до начала анализа.
3. 3) разные этапы подготовки данных после начала анализа.

Верный ответ: 2

7. Если хотим добавить новый объект к уже сформированным кластерам, тогда приступаем к

Ответы:

1. 1) кластеризации
2. 2) классификации

Верный ответ: 2

8. Если мы хотим распределить некий набор объектов по группам, тогда мы реализуем...

Ответы:

1. 1) группировку
2. 2) кластеризацию
3. 3) классификацию

Верный ответ: 2

9. Выбрать аналитические системы из следующего списка:

Ответы:

1. 1) Analitica
2. 2) Knime
3. 3) MongoDB
4. 4) MS SQL Server

Верный ответ: 1 и 2

10. Коллекция для не реляционной СУБД, это .... для реляционной СУБД

Ответы:

1. 1) строка
2. 2) столбец
3. 3) индекс
4. 4) таблица

Верный ответ: 4

## ***II. Описание шкалы оценивания***

*Оценка: 5*

*Нижний порог выполнения задания в процентах: 90*

*Описание характеристики выполнения знания: из 3 пунктов билета студент должен ответить на все. Допускают мелкие неточности в одном из пунктов билета.*

*Оценка: 4*

*Нижний порог выполнения задания в процентах: 70*

*Описание характеристики выполнения знания: Студент отвечает на 2 из 3 пунктов билета.*

*Оценка: 3*

*Нижний порог выполнения задания в процентах: 50*

*Описание характеристики выполнения знания: Студент отвечает частично на 2 из 3 пунктов билета.*

## ***III. Правила выставления итоговой оценки по курсу***

учитывается работа студента в течении всего семестра.