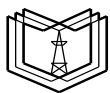


АКТУАЛИЗИРОВАНО  
Решением Ученого совета ИЦТЭ КГЭУ  
Протокол №7 от 19.03.2024



КГЭУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «КГЭУ»)

УТВЕРЖДАЮ

Директор Института цифровых  
технологий и экономики

\_\_\_\_\_ Э.И. Беляев

«30» мая 2023 г.

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

**Б1.В.ДЭ.01.01.01 Технологии больших данных**

---

Направление подготовки \_\_\_\_\_ 01.03.04 Прикладная математика

Направленность(и)  
(профиль(и)) \_\_\_\_\_ Технологии разработки цифровых систем и моделей

Квалификация \_\_\_\_\_ Бакалавр

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч.степень, уч.звание	ФИО разработчика
ЦСМ	к.т.н., доцент	Николаева С.Г.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	Цифровые системы и модели	19.05.2023	№ 5	_____ Зав. каф., к.ф.-м. н., доцент Ю.Н. Смирнов
Согласована	Учебно-методический совет института ЦТЭ	30.05.2023	№ 7	_____ Директор, к.т.н., доц. Беляев Э.И.
Одобрена	Ученый совет института ЦТЭ	30.05.2023	№ 9	_____ Директор, к.т.н., доц. Беляев Э.И.

## 1. Цель, задачи и планируемые результаты обучения по дисциплине

Целью освоения дисциплины «Технологии больших данных» является изучение студентами технологий и принципов обработки больших данных, практическая работа с современным инструментарием, таким, как фреймворки Hadoop и Hive, модель распределенных вычислений MapReduce.

Задачами дисциплины являются:

- ознакомление с возможностями технологий распределенного хранилища, спецификой хранения и обработки больших данных;
- освоение принципов разработки цифровых решений в области управления большими данными;
- развитие способности применять современные научные методы исследования и обработки информационных массивов;
- получение навыков установки и работы с платформой Hadoop, СУБД Hive;
- формирование практических навыков разработки приложений с помощью технологии MapReduce;
- изучение технологии цифровых решений на основе СУБД NoSQL-типа.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
ПК-1 Способен предлагать к внедрению актуальные цифровые решения задач предприятия	ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия
	ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия

## 2. Место дисциплины в структуре ОП

Предшествующие дисциплины (модули), практики, НИР, др.:

- Базы данных
- База данных приложений
- Производственная практика

Последующие дисциплины (модули), практики, НИР, др.:

- Технологии цифровых систем и моделей (параллельное изучение)
- Технологии внедрения цифровых систем и моделей (параллельное изучение)
- Выполнение ВКР

## 3. Структура и содержание дисциплины

### 3.1. Структура дисциплины

Для очной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	4	144	144
КОНТАКТНАЯ РАБОТА*	-	67	67
АУДИТОРНАЯ РАБОТА	1,44	52	52
Лекции	0,5	18	18
Практические (семинарские) занятия	0	0	0
Лабораторные работы	0,94	34	34
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	2,56	92	92
Проработка учебного материала	1,56	56	56
Курсовой проект	0	0	0
Курсовая работа	0	0	0
Подготовка к промежуточной аттестации	1	36	36
Промежуточная аттестация:			-
			Э

### 3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Введение в большие данные. Жизненный цикл аналитики данных	12	2			10	ТК1	ПК-1.1, 3 ПК-1.2, 3
Раздел 2. Технология Hadoop. СУБД Hive. Вычисления в MapReduce	37	6	16		15	ТК2	ПК-1.1, ЗУВ ПК-1.2, ЗУВ
Раздел 3. Масштабирование и многоуровневое хранение данных	29	6	8		15	ТК3	ПК-1.1, ЗВ ПК-1.2, ЗВ
Раздел 4. Визуализация данных и результатов анализа	30	4	10		16	ТК4	ПК-1.1, ЗУВ ПК-1.2, ЗУВ
Экзамен	36				36	<b>ОМ</b>	ПК-1.1, ЗУВ ПК-1.2, ЗУВ
<b>Итого за 7 семестр</b>	<b>144</b>	<b>18</b>	<b>34</b>		<b>92</b>		
<b>ИТОГО</b>	<b>144</b>	<b>18</b>	<b>34</b>		<b>92</b>		

### 3.3. Содержание дисциплины

Раздел 1. Введение в большие данные. Жизненный цикл аналитики данных.

Тема 1.1. Введение в большие данные.

Понятие Big Data. Причины появления, свойства, специфика работы с большими данными. Основные типы данных. Задачи, решаемые с помощью технологий Big Data.

Тема 1.2. Жизненный цикл аналитики данных.

Аналитика данных. Сбор, структурирование и анализ на основе Business Intelligence. Инструментарий BI. Оперативная аналитическая обработка (OLAP-технология).

Раздел 2. Технология Hadoop. Вычисления в MapReduce.

Тема 2.1. Технология Hadoop.

История и архитектура Hadoop. Файловая система HDFS. Движки MapReduce, Spark, Tez. Реляционные СУБД Hive, Shark, Spark SQL. Нереляционные БД: HBase.

Тема 2.2. Вычисления в MapReduce.

Модель распределенных вычислений MapReduce. Реализация для параллельной обработки больших массивов данных.

Раздел 3. Масштабирование и многоуровневое хранение данных.

Тема 3.1. Концепция СУБД NoSQL. Сравнение с СУБД SQL-типа. Виды, особенности, область применения. Масштабируемость. Репликация. CAP теорема. СУБД MongoDB.

Раздел 4. Визуализация данных и результатов анализа.

Тема 4.1. Типы, задачи, виды визуализации. Интерактивный сторителлинг, дашборды. Язык R, характеристики и возможности. Дедупликация данных.

### **3.4. Тематический план практических занятий**

Данный вид работы не предусмотрен учебным планом.

### **3.5. Тематический план лабораторных работ**

1. Установка и настройка Hadoop для Windows.
2. Работа в среде файловой системы HDFS.
3. Реализация вычислительного алгоритма в MapReduce.
4. Установка, настройка и работа с данными в Apache Hive.
5. Установка СУБД MongoDB. Работа с тестовой базой данных.
6. Репликация и шардинг данных в СУБД MongoDB.
7. Язык R. Изучение возможностей в среде RStudio.
8. Анализ временных рядов и прогнозирование в R.

### **3.6. Курсовой проект /курсовая работа**

Данный вид работы не предусмотрен учебным планом.

## **4. Оценивание результатов обучения**

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-1	ПК-1.1	знать:				
		терминологию, принципы организации, типы больших данных, круг решаемых задач, технологии сбора и анализа больших массивов данных, отличия реляционных баз данных от нереляционных	Глубокое знание терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных	Достаточно высокие знания терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных	Неуверенно знает терминологию, принципы организации, типы больших данных, круг решаемых задач сбора и анализа больших массивов данных, отличия реляционных баз данных от нереляционных	Знания терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных ниже минимального уровня
		уметь:				
		готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio	Уверенно и грамотно умеет готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio	На высоком уровне умеет готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio, немного	Низкий уровень умений подготовки аналитического материала для оценивания действий в различных областях для принятия решений, составления прогноза показателей деятельности предприятия в среде RStudio	Не может правильно готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio

			ошибается		
		<b>владеть:</b>			
	методами обработки больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	В совершенстве владеет методами обработки больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	Владеет в целом методами обработки больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	Низкая степень владения методами обработки больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	Не владеет методами обработки больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R
		<b>знать:</b>			
	современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	В полном объеме знает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	Практически полностью знает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений, допускает неточности в описании	Плохо или частично описывает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	Не может раскрыть терминологию и описать современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений
		<b>уметь:</b>			
	работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия	Умеет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия, в работе не ошибается	Умет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия, делает негрубые ошибки	Низкий уровень умений работы с большими данными в среде СУБД NoSQL-типа, разработки и адаптации программных компонентов обработки данных для решения задач предприятия, допускает серьезные ошибки	Не умеет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия
ПК-1.2					

		<b>владеть:</b>			
	навыками разработки программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	Уверенно владеет навыками разработки программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	Достаточно хорошо владеет навыками разработки программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	Плохо владеет навыками разработки программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce, часто ошибается	Не владеет навыками разработки программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce

Оценочные материалы для проведения текущего контроля и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины.

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика.

## **5. Учебно-методическое и информационное обеспечение дисциплины**

### **5.1. Учебно-методическое обеспечение**

#### **5.1.1. Основная литература**

1. Анализ больших данных : учебное пособие / И.Б. Тесленко, В.Е. Крылов, А.М. Губернаторов [и др.]. — Москва : КноРус, 2023. — 295 с. — ISBN 978-5-406-10550-4. — URL: <https://book.ru/book/950469>. — Текст : электронный.

2. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман, Дж. Ульман ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2016. - 500 с. - URL: <http://new.ibooks.ru/bookshelf/364297>. - ISBN 978-5-97060-190-7 : Б. ц. - Текст : электронный.

3. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 188 с. — ISBN 978-5-507-46866-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/322664>



### 5.1.2. Дополнительная литература

1. Базы данных в высокопроизводительных информационных системах : учебное пособие / составитель Е. И. Николаев. — Ставрополь : СКФУ, 2016. — 163 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/155218>

2. Григорьев, Ю. А. Реляционные базы данных и системы NoSQL : учебное пособие / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. — Благовещенск : АмГУ, 2018. — 424 с. — ISBN 978-5-93493-308-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/156492>

3. Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/145102>

4. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/182452>

## 5.2. Информационное обеспечение

### 5.2.1. Электронные и интернет-ресурсы

1. Электронно-библиотечная система «Лань», <https://e.lanbook.com/>
2. Портал «Открытое образование», <http://npoed.ru>
3. Единое окно доступа к образовательным ресурсам, <http://window.edu.ru>

### 5.2.2. Профессиональные базы данных / Информационно-справочные системы

1. Российский портал поддержки PostgreSQL, [postgrespro.ru](http://postgrespro.ru)
1. Научная электронная библиотека, <http://elibrary.ru>
2. Российская государственная библиотека, <http://www.rsl.ru>
3. Образовательный портал, <http://www.ucheba.com>

### 5.2.3. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№	Наименование программного	Способ	Реквизиты подтверждающих
---	---------------------------	--------	--------------------------

п/п	обеспечения	распространения (лицензионное/сво бодно)	документов
1	1. Операционная система Windows 7 Профессиональная	лицензионное	Договор ПО ЛИЦ № 0000/20, лицензиар – ЗАО «ТаксНет Сервис»
2	2. Office Professional Plus 2007 Russian OLP NL	лицензионное	Договор № 225/10, лицензиар - ЗАО «СофтЛайнТрейд»
3	LMS Moodle	свободно	-
4	Браузер Chrome	свободно	-
5	Java Development Kit	свободно	-
6	Apache Hadoop	свободно	-
7	СУБД MongoDB	свободно	-

## 6. Материально-техническое обеспечение дисциплины

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
Лекции	Учебная аудитория для проведения занятий лекционного типа	Специализированная учебная мебель, технические средства обучения, служащие для представления учебной информации большой аудитории (мультимедийный проектор, компьютер (ноутбук), экран), демонстрационное оборудование, учебно- наглядные пособия
Промежуточная аттестация	Учебная аудитория для проведения занятий семинарского типа, групповых и индивиду-альных консультаций, текущего контроля и промежуточной аттестации Д-424, Д-427	Специализированная учебная мебель, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран) и др.
Лабораторные работы	Учебно-научная лаборатория информационно- математического моделирования Д-424	Специализированное лабораторное оборудование по профилю лаборатории
	Компьютерный класс с выходом в Интернет Д-427	Специализированная учебная мебель, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран), лицензионное программное обеспечение
	Компьютерный класс с выходом в Интернет Д-427, Д- 424	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультиме- дийный проектор, компьютер (ноутбук), экран), видеокамеры, программное обеспечение
Самостоятельная работа	Компьютерный класс с выходом в Интернет Д-427	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультиме- дийный проектор, компьютер (ноутбук), экран), видеокамеры, программное

		обеспечение
	Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, экран, мультимедийный проектор, программное обеспечение

## **7. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов**

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета [www//kgeu.ru](http://www//kgeu.ru). Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;

- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;

- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;

- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;

- обеспечивается необходимый уровень освещенности помещений;

- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

## **8. Методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.**

Методическое обеспечение процесса воспитания обучающихся выступает одним из определяющих факторов высокого качества образования. Преподаватель вуза, демонстрируя высокий профессионализм, эрудицию, четкую гражданскую позицию, самодисциплину, творческий подход в решении профессиональных задач, в ходе образовательного процесса способствует формированию гармоничной личности.

При реализации дисциплины преподаватель может использовать следующие методы воспитательной работы:

- методы формирования сознания личности (беседа, диспут, внушение, инструктаж, контроль, объяснение, пример, самоконтроль, рассказ, совет, убеждение и др.);

- методы организации деятельности и формирования опыта поведения (задание, общественное мнение, педагогическое требование, поручение, приучение, создание воспитывающих ситуаций, тренинг, упражнение, и др.);

- методы мотивации деятельности и поведения (одобрение, поощрение социальной активности, порицание, создание ситуаций успеха, создание ситуаций для эмоционально-нравственных переживаний, соревнование и др.)

При реализации дисциплины преподаватель должен учитывать следующие направления воспитательной деятельности:

*Гражданское и патриотическое воспитание:*

- формирование у обучающихся целостного мировоззрения, российской идентичности, уважения к своей семье, обществу, государству, принятым в семье и обществе духовно-нравственным и социокультурным ценностям, к национальному, культурному и историческому наследию, формирование стремления к его сохранению и развитию;

- формирование у обучающихся активной гражданской позиции, основанной на традиционных культурных, духовных и нравственных ценностях российского общества, для повышения способности ответственно реализовывать свои конституционные права и обязанности;

- развитие правовой и политической культуры обучающихся, расширение конструктивного участия в принятии решений, затрагивающих их права и интересы, в том числе в различных формах самоорганизации, самоуправления, общественно-значимой деятельности;

- формирование мотивов, нравственных и смысловых установок личности, позволяющих противостоять экстремизму, ксенофобии, дискриминации по социальным, религиозным, расовым, национальным признакам, межэтнической и межконфессиональной нетерпимости, другим негативным социальным явлениям.

*Духовно-нравственное воспитание:*

- воспитание чувства достоинства, чести и честности, совестливости, уважения к родителям, учителям, людям старшего поколения;

- формирование принципов коллективизма и солидарности, духа милосердия и сострадания, привычки заботиться о людях, находящихся в трудной жизненной ситуации;

- формирование солидарности и чувства социальной ответственности по отношению к людям с ограниченными возможностями здоровья, преодоление психологических барьеров по отношению к людям с ограниченными возможностями;

- формирование эмоционально насыщенного и духовно возвышенного отношения к миру, способности и умения передавать другим свой эстетический опыт.

*Культурно-просветительское воспитание:*

- формирование эстетической картины мира;

- формирование уважения к культурным ценностям родного города, края, страны;

- повышение познавательной активности обучающихся.

*Научно-образовательное воспитание:*

- формирование у обучающихся научного мировоззрения;

- формирование умения получать знания;

- формирование навыков анализа и синтеза информации, в том числе в профессиональной области.

## Вносимые изменения и утверждения на новый учебный год

№ П/П	№ раздела внесения изменений	Дата внесения изменений	Содержание изменений	«Согласовано» Зав. каф. реализующей дисциплину	«Согласовано» председатель УМК института (факультета), в состав которого входит выпускающая
1	2	3	4	5	6
1					
2					
3					

*Приложение к рабочей  
программе дисциплины*



**КГЭУ**

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «КГЭУ»)**

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
по дисциплине**

**Б1.В.ДЭ.01.01.01 Технологии больших данных**

---

Направление подготовки

01.03.04 Прикладная математика

Квалификация

Бакалавр

г. Казань, 2023

Оценочные материалы по дисциплине «Технологии больших данных» предназначены для оценивания результатов обучения на соответствие индикаторам достижения компетенций.

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля (ТК) и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

## 1. Технологическая карта

Семестр 7

Наименование раздела	Формы и вид контроля	Рейтинговые показатели							
		I текущий контроль	Дополнительные баллы к ТК1	II текущий контроль	Дополнительные баллы к ТК2	III и VI текущий контроль	Дополнительные баллы к ТК3	Итого	Промежуточная аттестация
<b>Раздел 1. Введение в большие данные. Жизненный цикл аналитики данных</b>	<b>ТК1</b>	<b>15</b>	<b>0-15</b>					<b>15-30</b>	<b>15-30</b>
Коллоквиум		7							
Защита лабораторной работы		8							
<b>Раздел 2. Технология Hadoop. СУБД Hive. Вычисления в MapReduce</b>	<b>ТК2</b>			<b>15</b>	<b>0-15</b>			<b>15-30</b>	<b>15-30</b>
Коллоквиум				7					
Защита лабораторной работы				8					
<b>Раздел 3. Масштабирование и многоуровневое хранение данных</b>	<b>ТК3</b>					<b>10</b>	<b>0-10</b>	<b>10-30</b>	<b>10-30</b>
Коллоквиум						6			
Защита лабораторной работы						4			
<b>Раздел 4. Визуализация данных и результатов анализа</b>	<b>ТК4</b>					<b>15</b>	<b>0-15</b>	<b>15-30</b>	<b>15-30</b>
Коллоквиум						7			
Защита лабораторной работы						8			
<b>Промежуточная аттестация (экзамен)</b>	<b>ОМ</b>								<b>0-45</b>
Задание промежуточной аттестации (тест)									<b>0-45</b>

## 2. Оценочные материалы текущего контроля и промежуточной аттестации

Шкала оценки результатов обучения по дисциплине:

Код	Код	Заплани-	Уровень сформированности
-----	-----	----------	--------------------------



компетенции	индикатора компетенции	рованные результаты обучения по дисциплине	индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
зачтено			не зачтено			
ПК-1	ПК-1.1	знать:				
		терминологию, принципы организации, типы больших данных, круг решаемых задач, технологии сбора и анализа больших массивов данных, отличия реляционных баз данных от нереляционных	Глубокое знание терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных	Достаточно высокие знания терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных. Допускает незначительные ошибки	Неуверенно знает терминологию, принципы организации, типы больших данных, круг решаемых задач, технологии сбора и анализа больших массивов данных, отличия реляционных баз данных от нереляционных	Знания терминологии, принципов организации, типов больших данных, круга решаемых задач, технологий сбора и анализа больших массивов данных, отличий реляционных баз данных от нереляционных ниже минимального уровня
		уметь:				
		готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio	Уверенно и грамотно умеет готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio	На высоком уровне умеет готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio, немного ошибается	Низкий уровень умений подготовки аналитического материала для оценивания действий в различных областях для принятия решений, составления прогноза показателей деятельности предприятия в среде RStudio	Не может правильно готовить аналитический материал для оценивания действий в различных областях для принятия решений, составлять прогноз показателей деятельности предприятия в среде RStudio
		владеть:				
		методами обработки больших данных с помощью	В совершенстве владеет методами обработки	Владеет в целом методами обработки больших	Низкая степень владения методами обработки	Не владеет методами обработки больших данных с

		средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	больших данных с помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R	помощью средств и возможностей Hadoop, технологии MapReduce, файловой системы HDFS, языка R
ПК-1.2	знать:					
	современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	В полном объеме знает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	Практически полностью знает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений, допускает неточности в описании	Плохо или частично описывает современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	Не может раскрыть терминологию и описать современные программные средства и специфику работы с Big Data, технологии визуализации данных и результатов анализа, тенденции развития технологий исследования данных в сфере цифровых решений	
	уметь:					
	работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия	Умеет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия, в работе не ошибается	Умет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия, делает негрубые ошибки	Низкий уровень умений работы с большими данными в среде СУБД NoSQL-типа, разработки и адаптации программных компонентов обработки данных для решения задач предприятия, допускает серьезные ошибки	Не умеет работать с большими данными в среде СУБД NoSQL-типа, разрабатывать и адаптировать программные компоненты обработки данных для решения задач предприятия	
владеть:						
навыками разработки программного обеспечения на	Уверенно владеет навыками разработки программного	Достаточно хорошо владеет навыками разработки	Плохо владеет навыками разработки программного обеспечения	Не владеет навыками разработки программного		

		языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	программного обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce	на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce, часто ошибается	обеспечения на языке R, методиками проведения высокопроизводительных вычислений на основе фреймворка Hadoop и технологии MapReduce
--	--	---	--	---	---	--

Оценка **«отлично»** выставляется, если в течение семестра успешно выполнены лабораторные работы и пройден текущий контроль; студент глубоко и прочно усвоил программный материал, дает безошибочные ответы на тестовые задания.

Оценка **«хорошо»** выставляется, если в течение семестра успешно выполнены лабораторные работы и пройден текущий контроль; студент твердо знает материал, допускает несущественные ошибки в ответе на тестовые задания.

Оценка **«удовлетворительно»** выставляется, если в течение семестра выполнены лабораторные работы и пройден текущий контроль; студент имеет знания основного материала, но не усвоил его деталей, затрудняется при выполнении тестовых заданий.

Оценка **«неудовлетворительно»** выставляется за низкий уровень и/или неполное выполнение лабораторных работ и посредственные результаты при прохождении текущего контроля; студент не знает значительной части программного материала, с большими затруднениями выполняет тестовые задания.

### 3. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Описание оценочного средства
Коллоквиум (К)	Средство контроля усвоения учебного материала темы/раздела или разделов дисциплины, организованное как учебное занятие в виде собеседования преподавателя с обучающимися	Вопросы по темам / разделам дисциплины
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов. Оформление отчета, защита результатов лабораторной работы по отчету	Перечень заданий и вопросов для защиты лабораторной работы, перечень требований к отчету
Тест (Тест)	Система стандартизированных заданий, позволяющая	Комплект тестовых

автоматизировать процедуру измерения уровня знаний и умений обучающегося	заданий
--	---------

#### **4. Перечень контрольных заданий или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины**

##### **Для текущего контроля ТК1:**

Проверяемая компетенция: ПК-1, ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия; ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия.

##### Вопросы

1. Какие задачи решают большие данные?
2. Что такое технология OLAP?
3. Какие виды масштабирования вы знаете? Дать характеристику.
4. Характеристика термина Business Intelligence применительно к информационным технологиям.
5. Какие виды работ включает в себя жизненный цикл аналитики данных?
6. Приведите примеры удачной реализации проектов в области больших данных в России и за рубежом.
7. Какие проблемы/сложности существуют в области хранения больших данных?
8. Приведите расшифровку и дайте технико-технологическое определение термина *Data mining*.
9. Какие типы данных с точки зрения структурированности вы знаете? Дать характеристику.
10. Какие понятия/определения объединяет в себе принцип трех «V» ?

##### **Для текущего контроля ТК2:**

Проверяемая компетенция: ПК-1, ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия; ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия.

Лабораторная работа. Установка и настройка Hadoop для Windows (задание к работе)

1. Предварительно: установить комплект разработки ПО Java Development Kit (JDK) и настроить переменную Java\_Home, которая указывает на корневую директорию JDK.
2. Скачать Hadoop версии не ниже 2.2 с официального сайта. Разместить в выбранную папку распакованный архив.
3. Выполнить настройку конфигурации Hadoop (отредактировать файлы HDFS, исполнителя задач, планировщика ресурсов с менеджером YARN).
4. Выполнить запуск Hadoop, предварительно отформатировав распределенную файловую систему.
5. Оформить отчет с представлением скриншотов действий.

##### Вопросы

1. Что такое Hadoop? Назначение, основные компоненты.
2. Опишите компонент хранения данных в Hadoop.
3. Назовите основные достоинства и недостатки MapReduce.
4. Есть ли разница между СУБД и Hadoop? Принципы хранения данных в РСУБД и Hadoop.
5. Принцип действия распределенной файловой системы HDFS (хранение и обработка данных).
6. Функции компонентов Hadoop Task Tracker и Data Node. Почему они необходимы на каждом компьютере?
7. Какие вы знаете режимы/варианты использования Hadoop?
8. Опишите назначение компонентов экосистемы Hadoop HIVE и HBase.
9. Для чего используется автономный режим Hadoop?
10. Какой узел отвечает за репликацию данных в Hadoop?

### Для текущего контроля ТКЗ:

Проверяемая компетенция: ПК-1, ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия; ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия.

Лабораторная работа. Репликация и шардинг данных в СУБД MongoDB (задание к работе)

1. Создать поддиректории для хранения данных узлов репликации.
2. Создать узлы репликации в соответствии с заданием (количество и номера портов).
3. Создать базу данных и коллекцию на первичном узле. Проконтролировать ее репликацию на вторичном узле. Включить
4. Проверить работоспособность системы, отключив первичный узел. Дополнить коллекцию данными. Включить отключенный узел, удостовериться, что он будет вторичным, и данные, добавленные ранее, реплицированы.
5. Создать поддиректории для хранения данных сегментов и сами сегменты. Создать БД и коллекцию, хранящуюся в одном сегменте.
6. Добавить данные в соответствии с заданием, удостовериться, что они распределены в разные сегменты по ключу.
7. Оформить отчет с представлением скриншотов результатов.

### Вопросы

1. Что такое репликация? Как она реализована в СУБД MongoDB?
2. Виды репликации в MongoDB.
3. Сколько процессов mongod необходимо открыть для создания репликации?
4. Для чего выполняется шардинг (сегментирование)? Какая команда запускает шардинг?
5. Как осуществляется сегментирование коллекции базы данных?
6. Отличительные особенности NoSQL. Когда целесообразно применять NoSQL-хранилища?
7. Назовите недостатки и преимущества нереляционных СУБД.
8. Какие свойства фигурируют в теореме CAP?

9. Что, согласно теореме CAP, возможно обеспечить в любой реализации распределенных вычислений?

10. Какие виды шардинга вы знаете? Дать характеристику.

#### **Для текущего контроля ТК4:**

Проверяемая компетенция: ПК-1, ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия; ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия.

Лабораторная работа. Анализ временных рядов и прогнозирование в R (задание к работе)

1. Выполнить подключение необходимых для работы пакетов.
2. Загрузить данные из временного ряда в соответствии с заданием.
3. Построить график временного ряда с графиками автокорреляционной и частной автокорреляционной функций.
4. Подобрать вручную порядок ARMA-модели (визуальный анализ графика).
5. Подобрать ARMA-модель автоматически.
6. Построить прогноз на указанное в задании число шагов. Показать прогноз на графике.
7. Оформить отчет с представлением скриншотов результатов.

#### **Вопросы**

1. Назовите отличие автокорреляционной и частной автокорреляционной функций.
2. Характеристика понятий *случайное блуждание* и *белый шум*.
3. Каким образом возможно выполнить сравнение моделей временного ряда по степени точности?
4. Какие пакеты из *tidyverse* необходимо подключить для работы с временными рядами?
5. Опишите базовые возможности пакета *lubridate*.
6. Какой вид имеет автокорреляционная функция случайного процесса  $X(t)$  ?
7. Какие виды визуализации данных вы знаете?
8. Что такое интерактивный сторителлинг? Дашбординг?
9. Для каких целей разработан язык R? Основные возможности языка R.
10. Что такое дедупликация данных?

#### **Для промежуточной аттестации: ОМ**

Проверяемая компетенция: ПК-1, ПК-1.1 Способен выбирать актуальные цифровые решения задач предприятия; ПК-1.2 Способен внедрять и сопровождать актуальные цифровые решения задач предприятия.

Тест (Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика и содержит 50 тестовых вопросов 4-х типов (закрытые, открытые тесты, тесты на упорядочение, на установление соответствия) на каждую компетенцию)

1. Выберите технологию потоковой обработки событий в режиме реального времени
  - MapReduce
  - Apache Hadoop
  - Spark Streaming
  - Apache Kafka
2. Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк
  - Scikit-learn
  - TensorFlow
  - Flask
  - PyTorch
3. Apache NiFi используется для
  - визуализации результатов аналитики
  - эффективного хранения больших данных
  - оптимизации SQL-запросов к DWH
  - маршрутизации потоков Big Data и построения ETL-конвейеров
4. Повысить производительность Apache Kafka можно с помощью:
  - Увеличения плотности разделов на каждом брокере
  - Повышения коэффициента репликации
  - Замены HDD-дисков на SSD
  - Увеличения размера сообщений
5. Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит
  - Apache Hadoop
  - Apache Spark
  - Apache AirFlow
  - Apache Kafka
6. Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов
  - Можно
  - Нельзя
7. Формат Parquet считается
  - полуструктурированным
  - неструктурированным
  - строковым
  - колоночным (столбцовым)
8. Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON особенно подходит СУБД
  - Cassandra
  - Elasticsearch
  - HBase
  - Hive
9. Для машинного обучения подходят данные
  - Предварительно подготовленные, очищенные от ошибок, пропусков и

выбросов, а также нормализованные и представленные в виде числовых векторов

- Числовые типа int
- Любых форматов в цифровом виде
- Бинарные

10. Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью

- Apache Kafka
- Apache AirFlow
- Apache Hadoop
- Apache Hive

11. Какие из следующих технологий СУБД не используют принцип MapReduce:

- Hadoop
- Cassandra
- Redis

- HDInsight

12. Отметьте те из вариантов, в которых данные структурированы:

- данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word
- библиотека фильмов, представленных в формате mpreg4 на одном жестком диске
- таблица с ежедневными показаниями температуры помещения за год в файле формата csv
- текст педагогической поэмы А.С. Макаренко, представленный в формате PDF

13. Компания, проводящая социологические опросы получает анкеты от волонтеров, непосредственно опрашивающих респондентов. При каких условиях разумна постановка задачи цензурирования?

- часть анкет пришла в негодность, что не позволяет считать информацию с них со 100% уверенностью
- многие анкеты заполнены не полностью
- стало известно, что волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет
- от заказчика поступило требование уничтожить часть анкет, содержащих информацию о руководителях страны

14. В чём состоит свойство расширяемости записей СУБД?

- СУБД не имеет чёткой структуры, поэтому любую запись можно расширить



- повышение отказоустойчивости системы при добавлении новых записей в СУБД
- в любую таблицу СУБД можно добавить новую колонку, предварительно изменив структуру этой таблицы
- СУБД имеет чёткую, но расширяемую структуру, в каждую запись можно добавить новую колонку, также как и узнать значение любой записи по добавленной колонке

15. Отметьте причины создания NoSQL баз данных:

- высокая стоимость лицензий RDBMS
- высокая стоимость горизонтальной масштабируемости RDBMS при сохранении требования высокой доступности
- недостаточная гибкость языка запросов SQL
- невозможность хранить большие объёмы данных