



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

АКТУАЛИЗИРОВАНО  
Решением Ученого совета ИЦТЭ КГУ  
Протокол №7 от 19.03.2024

УТВЕРЖДАЮ  
Директор Института цифровых  
технологий и экономики

\_\_\_\_\_ Э.И. Беляев  
«30» мая 2023 г.

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

**Б1.О.15.04 Инжиниринг больших данных**

*(Код и наименование дисциплины в соответствии с РУП)*

Направление подготовки **09.03.01 Информатика и вычислительная техника**  
*(Код и наименование направления подготовки)*

Направленность(и)  
(профиль(и)) \_\_\_\_\_  
*(Наименование направленности (профиля) образовательной программы)*

Квалификация \_\_\_\_\_  
*(Бакалавр / Магистр)*

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч.степень, уч.звание	ФИО разработчика
ИТИС	Доцент, к.т.н, доцент	Хамитов Р.М.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	кафедра ИТИС	27.04.2023	3	_____ Зав.каф.,к.ф-м.н., доц. Соловьев С. А.
Согласована	кафедра ИТИС	27.04.2023	3	_____ Зав.каф.,к.ф-м.н., доц. Соловьев С. А.
Согласована	Учебно-методический совет института ИЦТЭ	30.05.2023	7	_____ Директор, к.т.н., доц. Беляев Э.И.
Одобрена	Ученый совет института ИЦТЭ	30.05.2023	9	_____ Директор, к.т.н., доц. Беляев Э.И.

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

*(Цель и задачи освоения дисциплины, соответствующие цели ОП)*

Целью освоения дисциплины "Инжиниринг больших данных" является формирование у студентов теоретических и практических знаний о методах сбора, извлечения и обработки данных, методах построения систем хранения данных, обучении технологии функционирования платформ распределенной обработки больших наборов данных.

Задачами дисциплины являются:

- формирование знаний и навыков работы с данными, представленными в различных форматах;
- обучение методам подготовки данных для систем машинного обучения;
- формирование знаний в области систем хранения данных.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности;	ОПК-2.2 Владеет навыками применения цифровых технологий для решения задач профессиональной деятельности с учетом основных требований информационной безопасности

## 2. Место дисциплины в структуре ОП

Предшествующие дисциплины (модули), практики, НИР, др Информационные технологии, Информационная безопасность, Базы данных.

Последующие дисциплины (модули), практики, НИР, др. Разработка приложений искусственного интеллекта, Цифровые системы автоматизации технологических процессов и интернет-вещей, Проектный практикум, Инженерия цифровых двойников и систем, Технологии искусственного интеллекта, База данных и технологии распределенного реестра, Нейросетевые технологии в ИСУ, База данных информационных систем управления, Облачные технологии, Сквозные цифровые технологии, Учебная практика (ознакомительная), Производственная практика (проектная), Производственная практика (технологическая (проектно-технологическая), Производственная практика (преддипломная), Выполнение и защита выпускной квалификационной работы.

### 3. Структура и содержание дисциплины

#### 3.1. Структура дисциплины

Для очной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			6
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	4	144	144
КОНТАКТНАЯ РАБОТА*	-	61	61
АУДИТОРНАЯ РАБОТА	1.28	46	46
Лекции	0.45	16	16
Практические (семинарские) занятия	-	0	0
Лабораторные работы	0.83	30	30
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	2.72	98	98
Проработка учебного материала	1.72	62	62
Курсовой проект	-	0	0
Курсовая работа	-	0	0
Подготовка к промежуточной аттестации	1	36	36
Промежуточная аттестация:			Э
			-

Для заочной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Курс/Сессия
			4/7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	4	144	144
КОНТАКТНАЯ РАБОТА	-	41	41
АУДИТОРНАЯ РАБОТА	0,56	20	20
Лекции	0,22	8	8
Практические (семинарские) занятия	-	0	0
Лабораторные работы	0,33	12	12
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	3,44	124	124
Проработка учебного материала	3,19	115	115
Курсовой проект	-	0	0
Курсовая работа	-	0	0
Подготовка к промежуточной аттестации	0,25	9	9
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 Работа с данными в Python.		4	8		20	ТК1	ОПК-2.3, ОПК-2.У, ОПК-2.В
Раздел 2 Подготовка данных для систем машинного обучения.		6	12		20	ТК2	ОПК-2.3, ОПК-2.У, ОПК-2.В
Раздел 3 Параллельная и распределенная обработка данных.		6	12		20	ТК3	ОПК-2.3, ОПК-2.У, ОПК-2.В
Экзамен	36				36	<b>ОМ 3</b>	<b>ОПК-2.3, ОПК-2.У, ОПК-2.В</b>
<b>Итого за 6 семестр</b>	<b>144</b>	<b>16</b>	<b>30</b>	<b>0</b>	<b>98</b>		
<b>ИТОГО</b>	<b>144</b>	<b>16</b>	<b>30</b>	<b>0</b>	<b>98</b>		

### 3.3. Содержание дисциплины

Раздел 1. Работа с данными в Python.

Тема 1.1. Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python.

Тема 1.2. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний.

Раздел 2 Подготовка данных для систем машинного обучения.

Тема 2.1. Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей.

Тема 2.2. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.

Тема 2.3. Общедоступные платформы для хранения данных. Подход Data-Centric AI.

Раздел 3 Параллельная и распределенная обработка данных.

Тема 3.1. Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, Распределенная файловая система HDFS. Парадигма MapReduce.

Тема 3.2. Решение задач с помощью MapReduce. Алгоритмы на графах в MapReduce. Pig и Hive. NoSQL базы данных: HBase и Cassandra. YARN. MapReduce 2.0

Тема 3.3. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark

DataFrame. Использование SQL в Spark DataFrame

### 3.4. Тематический план практических занятий

*«Данный вид работы не предусмотрен учебным планом».*

### 3.5. Тематический план лабораторных работ

Раздел 1. Работа с данными в Python.

Тема 1.1. Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python.

Тема 1.2. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний.

Раздел 2 Подготовка данных для систем машинного обучения.

Тема 2.1. Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей.

Тема 2.2. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.

Тема 2.3. Общедоступные платформы для хранения данных. Подход Data-Centric AI.

Раздел 3 Параллельная и распределенная обработка данных.

Тема 3.1. Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, Распределенная файловая система HDFS. Парадигма MapReduce.

Тема 3.2. Решение задач с помощью MapReduce. Алгоритмы на графах в MapReduce. Pig и Hive. NoSQL базы данных: HBase и Cassandra. YARN. MapReduce 2.0

Тема 3.3. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame

### 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
ОПК-2	ОПК-2.2	<b>знать:</b>				
		принципы и методы хранения наборов данных различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта	Уровень знаний в объеме, соответствует вующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствует вующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		существующие и перспективные структуры центров обработки данных	Уровень знаний в объеме, соответствует вующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствует вующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		<b>уметь:</b>				
		применять принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения
	применять принципы и методы хранения	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые	

		наборов данных различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков и области искусственного интеллекта			оценивать	умения
		владеть:				
		навыками применения цифровых технологий для решения задач профессиональной деятельности с учетом основных требований информационной безопасности	Продемонстрированы навыки	Сформированы базовые навыки	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки, допущены значительные ошибки.

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. 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>. — Режим доступа: для авториз. пользователей.

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>.

#### 5.1.2. Дополнительная литература

1. Харрисон М. Как устроен Python. Гид для разработчиков, программистов и интересующихся / М. Харрисон. - Санкт-Петербург : Питер, 2019. - 272 с. - ISBN 978-5-4461-0906-7. - URL: <https://ibooks.ru/bookshelf/359217/reading> - Текст: электронный.

2. Хахаев, И. А. Практикум по алгоритмизации и программированию на Python : учебное пособие / И. А. Хахаев. — 2-е изд. — Москва : ИНТУИТ, 2016. — 178 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/100377>. — Режим доступа: для авториз. пользователей.

3. Наместников, А. М. Базы данных. Практический курс : учебное пособие : в 2 частях / А. М. Наместников. — Ульяновск : УлГТУ, 2017 — Часть 1 : Объектно-реляционные базы данных на примере PostgreSQL 9.5 — 2017. — 113 с. — ISBN 978-5-9795-1743-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165100>.

4. Дадян, Эдуард Григорьевич. Данные: хранение и обработка : учебник / Э. Г. Дадян. - Москва : ИНФРА-М, 2021. - 205 с. - URL: <http://new.ibooks.ru/bookshelf/377999>. - ISBN 978-5-16-016447-2 (print). - ISBN 978-5-16-107936-2 (online). - Текст : электронный.

5. Волк, В. К. Базы данных. Проектирование, программирование, управление и администрирование : учебник / В. К. Волк. - 3-е изд., стер. - Санкт-Петербург : Лань, 2022. - 242 с. - URL: <https://e.lanbook.com/book/193373>. - ISBN 978-5-8114-9368-5. - Текст : электронный.

## 5.2. Информационное обеспечение

### 5.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	<a href="https://e.lanbook.com/">https://e.lanbook.com/</a>
2	Электронно-библиотечная система «ibooks.ru»	<a href="https://ibooks.ru/">https://ibooks.ru/</a>
3	Электронно-библиотечная система «book.ru»	<a href="https://www.book.ru/">https://www.book.ru/</a>
4	Портал «Открытое образование»	<a href="https://npoed.ru">https://npoed.ru</a>
5	Российская национальная библиотека	<a href="https://nlr.ru/">https://nlr.ru/</a>
6	КиберЛенинка	<a href="https://cyberleninka.ru">https://cyberleninka.ru</a>
7	Техническая библиотека	<a href="https://techlibrary.ru">https://techlibrary.ru</a>
8	Национальная электронная библиотека (НЭБ)	<a href="https://rusneb.ru/">https://rusneb.ru/</a>
9	Учебный курс Надоор. Система для обработки больших объемов данных	<a href="https://stepik.org/course/150/syllabus">https://stepik.org/course/150/syllabus</a>
10	Real Python Tutorials.	<a href="https://realpython.com/">https://realpython.com/</a>
11	Data Engineering: ETL, ELT, Data Pipeline, Data Warehouse, Data Lakes, Data Marts	<a href="https://ivan-shamaev.ru/data-engineering-etl-pipeline-data-warehouse-datalake/">https://ivan-shamaev.ru/data-engineering-etl-pipeline-data-warehouse-datalake/</a>
12	Data-Learn / data-engineering	<a href="https://github.com/Data-Learn/data-engineering/blob/master/DE%20-%20101%20Guide.md">https://github.com/Data-Learn/data-engineering/blob/master/DE%20-%20101%20Guide.md</a>
13	Spark SQL, DataFrames and Datasets Guide.	<a href="https://spark.apache.org/docs/latest/sql-programming-guide.html">https://spark.apache.org/docs/latest/sql-programming-guide.html</a>
14	Jake VanderPlas. Python Data Science Handbook.	<a href="https://jakevdp.github.io/PythonDataScienceHandbook/">https://jakevdp.github.io/PythonDataScienceHandbook/</a>
15	Jeroen Janssens. Data Science at the Command Line.	<a href="https://www.datascienceatthecommandline.com/">https://www.datascienceatthecommandline.com/</a>

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

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	Официальный интернет-портал правовой информации	<a href="http://pravo.gov.ru">http://pravo.gov.ru</a>	<a href="http://pravo.gov.ru">http://pravo.gov.ru</a>
2	Справочная правовая система «Консультант Плюс»	<a href="http://consultant.ru">http://consultant.ru</a>	<a href="http://consultant.ru">http://consultant.ru</a>

3	Справочно-правовая система по законодательству РФ	<a href="http://garant.ru">http://garant.ru</a>	<a href="http://garant.ru">http://garant.ru</a>
---	---	---	---

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

№ п/п	Наименование программного обеспечения	Описание	Реквизиты подтверждающих документов
1	Операционная система Microsoft Windows 10	Пользовательская операционная система	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
2	Microsoft Office 2019	Пакет офисных приложений	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
3	LMS Moodle	ПО для эффективного онлайн- взаимодействия преподавателя и студента	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
4	Браузер Chrome	Система поиска информации в сети интернет	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
5	google colab -	Веб - среда разработки для языка программирования Python. <a href="https://colab.research.google.com/">https://colab.research.google.com/</a>	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
6	Anaconda solution	Дистрибутив языков программирования Python и R, включающий набор популярных свободных библиотек, объединённых проблематиками науки о данных и машинного обучения. <a href="https://www.anaconda.com/">https://www.anaconda.com/</a>	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
7	Apache Spark	Система распределенной обработки данных <a href="https://spark.apache.org/">https://spark.apache.org/</a>	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.

## 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
Лабораторные работы	Учебная лаборатория программной инженерии, ауд. В-608	Специализированное лабораторное оборудование по профилю лаборатории программной инженерии, специализированная учебная мебель на 50 посадочных мест, 24 компьютера с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-610	Специализированная учебная мебель на 42 посадочных места, 17 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория информационной безопасности, ауд. В-615	Специализированное лабораторное оборудование по профилю лаборатории информационной безопасности, специализированная учебная мебель на 35 посадочных мест, 15 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-617	Специализированная учебная мебель на 24 посадочных места, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-619	Специализированная учебная мебель на 26 посадочных мест, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое

		лицензионное программное обеспечение
	Компьютерный класс, ауд. В-621	Специализированная учебная мебель на 35 посадочных мест, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория реинжиниринга и управления бизнес-процессами, ауд. В-623	Специализированное лабораторное оборудование по профилю лаборатории реинжиниринга и управления бизнес-процессами, специализированная учебная мебель на 34 посадочных места, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, ноутбук, экран), видеокамеры, необходимое лицензионное программное обеспечение
Самостоятельная работа	Компьютерный класс с возможностью выхода в Интернет и обеспечением доступа в ЭИОС В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный 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*



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.О.15.04 Инжиниринг больших данных**  
*(Наименование дисциплины в соответствии с учебным планом)*

---

Направление подготовки      09.03.01 Информатика и вычислительная техника  
*(Код и наименование направления подготовки)*

Квалификация      Бакалавр  
*(Бакалавр / Магистр)*

г. Казань, 2023

Оценочные материалы по дисциплине "Инжиниринг 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 6

Наименование раздела	Формы и вид контроля	Рейтинговые показатели							
		I текущий контроль	Дополнительные баллы к ТК1	II текущий контроль	Дополнительные баллы к ТК2	III текущий контроль	Дополнительные баллы к ТК3	Итого	Промежуточная аттестация
<b>Раздел 1 Работа с данными в Python.</b>	<b>ТК1</b>	<b>15</b>	<b>0-15</b>					<b>15-30</b>	<b>15-30</b>
Защита лабораторной работы		10							
Тест или письменный опрос		5							
<b>Раздел 2 Подготовка данных для систем машинного обучения.</b>	<b>ТК2</b>			<b>20</b>	<b>0-15</b>			<b>20-30</b>	<b>20-35</b>
Защита лабораторной работы				15					
Тест или письменный опрос				5					
<b>Раздел 3 Параллельная и распределенная обработка данных.</b>	<b>ТК3</b>					<b>20</b>	<b>0-15</b>	<b>20-30</b>	<b>20-35</b>
Защита лабораторной работы						15			
Тест или письменный опрос						5			
<b>Промежуточная аттестация (экзамен)</b>	<b>ОМ</b>								<b>0-45</b>
Тест или письменный опрос									45

## 2. Оценочные материалы текущего контроля и промежуточной аттестации

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
ОПК-2	ОПК-2.2	<b>знать:</b>					
		принципы и методы хранения наборов данных различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта	Уровень знаний в объеме, соответствует вующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствует вующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	
		существующие и перспективные структуры центров обработки данных	Уровень знаний в объеме, соответствует вующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствует вующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	
		<b>уметь:</b>					
		применять принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения	
применять принципы и методы хранения наборов данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения оценивать	Не сформированы базовые умения			

		различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков и области искусственного интеллекта				
		владеть:				
		навыками применения цифровых технологий для решения задач профессиональной деятельности с учетом основных требований информационной безопасности	Продемонстрированы навыки	Сформированы базовые навыки	Имеется минимальный набор навыков	Не продемонстрированы базовые навыки, допущены значительные ошибки.

Оценка **«отлично»** выставляется за выполнение лабораторных работ в семестре; тестовых заданий; умение применять принципы и методы хранения наборов данных различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков и области искусственного интеллекта; владение навыками применения цифровых технологий для решения задач профессиональной деятельности с учетом основных требований информационной безопасности;

Оценка **«хорошо»** выставляется за выполнение лабораторных работ в семестре; тестовых заданий; понимание принципов и методов хранения наборов данных различных типов на общедоступных платформах для обеспечения потребностей организаций-разработчиков и области искусственного интеллекта;

Оценка **«удовлетворительно»** выставляется за выполнение лабораторных работ в семестре и тестовых заданий;

Оценка **«неудовлетворительно»** выставляется за слабое и неполное выполнение лабораторных работ в семестре и тестовых заданий.

### 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
Опрос по разделам (темам)	Знание основных понятий темы/раздела/дисциплины	Перечень определений основных понятий темы/дисциплины
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов испытаний, измерений, эксперимента. 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 автоматизировать процедуру измерения уровня знаний и умений обучающегося	Комплект тестовых заданий

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

*Пример задания*

**Для текущего контроля ТК1:**

Проверяемая компетенция: ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности;

ОПК-2.2 Владеет навыками применения цифровых технологий для решения задач профессиональной деятельности с учетом основных требований информационной безопасности

**Отчет по лабораторной работе**

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 2 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в

электронном виде в виртуальной образовательной среде Moodle.

Темы лабораторных работ:

Тема 1.1. Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python.

Тема 1.2. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний.

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен верно, все тестовые программы проведены, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении тестовых отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки при выполнении тестирования программ - 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Максимальное количество баллов за все задания раздела – 15 баллов.

### Тест

<i>Вопрос</i>	<i>Варианты ответа</i>
<i>Какой логической операции не существует в математической логике?</i>	<i>логическое И (конъюнкция)</i>
	<i>логическое ИЛИ (дизъюнкция)</i>
	<b><i>логическое ДЛЯ (субтракция)</i></b>
	<i>логическое отрицание (инверсия)</i>
<i>Что является машино-ориентированным языком?</i>	<i>C++</i>
	<b><i>ассемблер</i></b>
	<i>Java</i>
	<i>Python</i>
<i>Какой из перечисленных языков скорее предметно-ориентированный нежели общего назначения?</i>	<i>C#</i>
	<b><i>SQL</i></b>
	<i>Java</i>
	<i>Python</i>
<i>Какого этапа нет в работе интерпретатора Python?</i>	<i>Инициализация</i>
	<b><i>Связывание</i></b>
	<i>Компиляция</i>
	<i>Выполнение</i>
<i>Какие циклы есть в Python кроме for(введите команду)?</i>	<b><i>while</i></b>

### **Для текущего контроля ТК2:**

Проверяемая компетенция: ПК-1 Способен проектировать программные приложения и создавать программные прототипы решения прикладных задач организационного управления и бизнес-процессов

ПК-1.1 Использует современные объектно-ориентированные, структурные языки программирования и языки бизнес-приложений

ПК-1.2 Применяет средства и методы проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов

#### **Отчет по лабораторной работе**

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 2 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

#### **Темы лабораторных работ:**

Тема 2.1. Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей.

Тема 2.2. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных.

Тема 2.3. Общедоступные платформы для хранения данных. Подход Data-Centric AI.

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен верно, все тестовые программы проведены, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении тестовых отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки при выполнении тестирования программ - 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Максимальное количество баллов за все задания раздела – 15 баллов.

## Тест

Вопрос	Варианты ответа
Как в библиотеке NumPy вычислить размеры массива (его измерения)?	<code>ndarray.ndim</code>
	<b><code>ndarray.shape</code></b>
	<code>ndarray.size</code>
	<code>ndarray.itemsize</code>
Как в библиотеке NumPy создать многомерный массив из списка?	<code>np.init([1, 2, 3])</code>
	<b><code>np.array([1, 2, 3])</code></b>
	<code>np.create([1, 2, 3])</code>
	<code>np.fill([1, 2, 3])</code>
Каких создать матрицу из нулей в NumPy?	<code>np.empty((3, 5))</code>
	<code>np.zeroInit((3, 5))</code>
	<b><code>np.zeros((3, 5))</code></b>
	<code>np.zero((3, 5))</code>
В NumPy матрицу можно создать различными способами. Выберите неправильный вариант.	<code>np.zeros()</code>
	<code>np.eye()</code>
	<b><code>np.full()</code></b>
	<code>np.ones()</code>
В NumPy матрицу можно создать различными способами. Выберите правильный вариант.	<code>np.full()</code>
	<b><code>np.empty()</code></b>
	<code>np.ear()</code>
	<code>np.twos()</code>

### Для текущего контроля ТКЗ:

Проверяемая компетенция: ПК-2 Способен осуществлять проектно-аналитические работы с использованием технологий больших данных и искусственного интеллекта

ПК-2.1 Способен планировать и организовывать аналитические работы с использованием технологий больших данных

ПК-2.2 Выполняет полный цикл решения задач с помощью машинного обучения и продвинутой аналитики

### Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 2 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Темы лабораторных работ:

Тема 3.1. Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для

распределенного хранения и обработки больших объемов данных: Apache Hadoop, Распределенная файловая система HDFS. Парадигма MapReduce. Тема 3.2. Решение задач с помощью MapReduce. Алгоритмы на графах в MapReduce. Pig и Hive. NoSQL базы данных: HBase и Cassandra. YARN. MapReduce 2.0

Тема 3.3. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен верно, все тестовые программы проведены, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении тестовых отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки при выполнении тестирования программ - 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Максимальное количество баллов за все задания раздела – 15 баллов.

Тест

<i>Вопрос</i>	<i>Варианты ответа</i>
<i>Выберите для чего не используется фреймворк ApacheSpark.</i>	обработка больших данных
	разработка потоковых приложений NRT
	<b>онлайн процессинг данных OLTP</b>
	работа со структурированными данными SQL
<i>Apache Spark используется во многих сферах. Выберите один верный вариант.</i>	онлайн процессинг данных OLTP
	<b>построение ETL конвейеров</b>
	приложения встраиваемых систем
	обработка данных менее 100 МБ
<i>Какое определение драйвера наиболее верное.</i>	контролирует статус работы
	распределяет ресурсы между приложениями
	<b>управляет через объект SparkContext выполнением задач</b>
	выполняет задачи(tasks) для вычисления
<i>Какое определение менеджера ресурсов наиболее верное.</i>	контролирует статус работы
	<b>распределяет ресурсы между приложениями</b>
	управляет через объект SparkContext выполнением задач
	выполняет задачи(tasks) для вычисления

**Для промежуточной аттестации:**  
**Тест**

Вопрос	Варианты ответа
Как в библиотеке NumPy вычислить размеры массива (его измерения)?	<code>ndarray.ndim</code>
	<b><code>ndarray.shape</code></b>
	<code>ndarray.size</code>
	<code>ndarray.itemsize</code>
Как в библиотеке NumPy создать многомерный массив из списка?	<code>np.init([1, 2, 3])</code>
	<b><code>np.array([1, 2, 3])</code></b>
	<code>np.create([1, 2, 3])</code>
	<code>np.fill([1, 2, 3])</code>
Каких создать матрицу из нулей в NumPy?	<code>np.empty((3, 5))</code>
	<code>np.zeroInit((3, 5))</code>
	<b><code>np.zeros((3, 5))</code></b>
	<code>np.zero((3, 5))</code>
В NumPy матрицу можно создать различными способами. Выберите неправильный вариант.	<code>np.zeros()</code>
	<code>np.eye()</code>
	<b><code>np.full()</code></b>
	<code>np.ones()</code>
В NumPy матрицу можно создать различными способами. Выберите правильный вариант.	<code>np.full()</code>
	<b><code>np.empty()</code></b>
	<code>np.ear()</code>
	<code>np.twos()</code>
Выберите для чего не используется фреймворк ApacheSpark.	обработка больших данных
	разработка потоковых приложений NRT
	<b>онлайн процессинг данных OLTP</b>
	работа со структурированными данными SQL
Apache Spark используется во многих сферах. Выберите один верный вариант.	онлайн процессинг данных OLTP
	<b>построение ETL конвейеров</b>
	приложения встраиваемых систем
	обработка данных менее 100 МБ
Какое определение драйвера наиболее верное.	контролирует статус работы
	распределяет ресурсы между приложениями
	<b>управляет через объект <code>SparkContext</code> выполнением задач</b>
	выполняет задачи(tasks) для вычисления
Какое определение менеджера ресурсов наиболее верное.	контролирует статус работы
	<b>распределяет ресурсы между приложениями</b>
	управляет через объект <code>SparkContext</code> выполнением задач
	выполняет задачи(tasks) для вычисления