

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ
Проректор по учебной работе,
качеству образования – первый
проректор
Хагуров Т.А.

подпись

«30» мая 2025

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.О.38 «Технологии обработки больших данных»

Направление подготовки 09.03.03 Прикладная информатика

Направленность (профиль) «Прикладная информатика в экономике»,
«Современные методы машинного
обучения и компьютерного зрения»

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины «Технологии обработки больших данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 09.03.03 Прикладная информатика.

Программу составил:

Калайдин Е.Н., д. ф.-м. н. профессор кафедры прикладной математики КубГУ



Рабочая программа дисциплины «Теория вероятности и математическая статистика» утверждена на заседании кафедры анализа данных и искусственного интеллекта

№9 от «6» мая 2025 г.

Заведующий кафедрой (разработчик) Письменский А.В., канд. физ.-мат. наук., доцент



подпись

Утверждена на заседании учебно-методической комиссии факультета компьютерных технологий и прикладной математики протокол

№4 от «23» мая 2025 г.

Председатель УМК факультета

А. В. Коваленко



подпись

Рецензенты:

Трофимов Виктор Маратович, доктор физико-математических наук, старший научный сотрудник, профессор Кафедры информационных систем и программирования ФГБОУ ВО «Кубанский государственный технологический университет»

Попова Елена Витальевна, доктор экономических наук, кандидат физико-математических наук, профессор, Заведующий кафедрой информационных систем Федерального государственного бюджетного образовательного учреждения высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина»

1 Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины

Формирование у студентов базовых знаний, умений и навыков в изучении основных понятий и принципов углубленного анализа данных, их взаимосвязи и развития, а также отвечающих им методов расчёта, используемых для анализа, моделирования и решения прикладных задач, связанных с большими объемами данных достаточных для освоения основной образовательной программы направления 09.03.03 Прикладная информатика; формирование составляющих частей общекультурных и профессиональных компетенций.

1.2 Задачи дисциплины

- подготовка специалистов, способных применять полученные знания для решения прикладных задач, владеющих достаточными знаниями основных теоретических положений курса «Технологии обработки больших данных»;
- формирование культуры мышления, способности к анализу, обобщению и восприятию информации, к постановке цели и выбору путей ее достижения;
- обеспечение математическим аппаратом естественнонаучных, общепрофессиональных и специальных дисциплин;
- формирование привычки к строгости в формулировке изложения материала, к логически непротиворечивой цепочке выводов и заключений;
- формирование способности производить анализ данных большого объема и на его основе предлагать пути решения поставленной задачи;
- развитие у студентов навыков самообразования.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Технологии обработки больших данных» относится к «К.М. Комплексные модули» Блока 1 «Дисциплины (модули)» учебного плана.

Необходимым требованием к «входным» знаниям, умениям и опыту деятельности обучающегося при освоении данной дисциплины является уверенное владение знаниями по предметам «Алгебра и аналитическая геометрия», «Математический анализ», «Дифференциальные и разностные уравнения», «Дискретная математика», «Численные методы», «Теория систем и системный анализ», «Теория вероятностей и математическая статистика», «Многомерный статистический анализ» и другие. Дисциплина призвана дать студентам математический аппарат, который будет использоваться в дальнейшем при выполнении студентами учебно-исследовательской работы

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
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	
ИОПК-2.1. Разрабатывает программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	Знает: Основные принципы работы современных информационных технологий и программных средств, в том числе отечественного производства и использовать их при решении задач профессиональной деятельности
	Умеет: Использовать существующие типовые решения и шаблоны проектирования программного обеспечения, использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
	Владеет (ТД): Разработка с использованием современных информационных технологий и программных средств и адаптацией существующих, изменение и согласование программного обеспечения, в том числе отечественного производства и использовать их при решении задач профессиональной деятельности
ОПК-7 Способен разрабатывать алгоритмы и программы, пригодные для практического применения	
ИОПК-7.1 Разрабатывает алгоритмы и программы, пригодные для практического применения	Знает: Методы и средства проектирования программного обеспечения при реализации математически сложных алгоритмов
	Умеет: применять методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов
	Владеет (ТД): Разработка алгоритмов и архитектуры программного обеспечения, пригодного для практического применения

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зач. ед. (72 часов), их распределение по видам работ представлено в таблице

Вид учебной работы		Всего часов	Семестры (часы)			
			7			
Контактная работа, в том числе:		52,2	52,2			
Аудиторные занятия (всего):		50	50			
Занятия лекционного типа		16	16			
Лабораторные занятия		34	34			
Занятия семинарского типа (семинары, практические занятия)						
Иная контактная работа:		2,2	2,2			
Контроль самостоятельной работы (КСР)		2	2			
Промежуточная аттестация (ИКР)		0,2	0,2			
Самостоятельная работа, в том числе:		19,8	19,8			
Проработка учебного (теоретического) материала		8	9,8			
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)		7	5			
Подготовка к текущему контролю		4,8	5			
Контроль:		-	-			
Подготовка к экзамену		-	-			
Общая трудоемкость	час.	72	72			
	в том числе контактная работа	52,2	52,2			
	зач. ед	2	2			

2.2 Структура дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 7 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	2	3	4	5	6	7
1.	Библиотека NumPy и Pandas	5			4	1
2.	Использование различных форматов файлов в задачах обработки данных.	5			4	1
3.	Взаимодействие с табличными данными в приложениях обработки данных.	5			4	1
4.	Визуализация данных	4			2	2
5.	Работа со строками в приложениях обработки данных	3			2	1
6.	Введение в обработку текста на естественном языке в задачах обработки данных	8	4		2	2
7.	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.	3			2	1
8.	Взаимодействие с базой данных в приложениях обработки данных.	5	2		2	1
9.	Параллельная обработка данных	10	4		4	2
10.	Библиотека Dask	7	2		4	1
11.	Обзор проблем обработки больших данных и вычисления общего назначения на GPU	10	4		4	2
ИТОГО по разделам дисциплины		65	16		34	15
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,2				
Подготовка к текущему контролю		4,8				
Общая трудоемкость по дисциплине		72				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Библиотека NumPy и Pandas	В рамках темы рассматривается технологический стек Python для обработки и анализа данных, возможности Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python. Организация массивов в NumPy: хранение данных,	Устный опрос. тестирование

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		создание массивов, принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры Маскирование и прихотливое индексирование в NumPy. В рамках темы рассматриваются возможности библиотеки Pandas. Организация Pandas DataFrame и организация индексации для DataFrame и Series; применение универсальных функций и работа с пустыми значениями в Pandas. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры. Рассматривается операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение».	
2.	Использование различных форматов файлов в задачах обработки данных.	В рамках темы рассматриваются принципы работы с файлами, файлы и операционные системы. Специфика текстовых и бинарных файлов. В рамках темы рассматривается задача сериализации и десериализации данных и использование различных форматов файлов для ее решения. Описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python. В рамках темы рассматриваются формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup. В рамках темы рассматривается проблематика форматов файлов для хранения и обработки больших данных. Форматы файлов NPY и HDF: общая характеристика, пример взаимодействия с данными этих форматов в Python.	Устный опрос. тестирование
3.	Взаимодействие с табличными данными в приложениях обработки данных.	В рамках темы рассматривается формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python. В рамках темы рассматриваются возможности использования Excel для внешних приложений обработки данных.	Устный опрос. тестирование

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		Взаимодействие с Excel из Python с помощью библиотеки XLWings: принципы работы и примеры использования.	
4.	Визуализация данных	В рамках темы рассматриваются основы работы с библиотекой matplotlib: организация системы координат, оформление осей, цвета и цветовые карты в matplotlib, стили линий и маркеры. Pyplot и объектно-ориентированный интерфейс matplotlib. Управление фигурами и создание множества графиков на одном рисунке. Различные типы графиков. В рамках темы рассматривается визуализация данных с помощью библиотеки Pandas: набор методов для построения графиков, реализованный в структурах Series и DataFrame. В рамках темы проводится введение в разведочный анализ данных: типы признаков, анализ распределений, анализ мер центральной тенденции и поиск выбросов, анализ взаимного распределения и парных корреляций. Проведение разведочного анализа данных с помощью библиотеки Seaborn	Устный опрос. тестирование
5.	Работа со строками в приложениях обработки данных	В рамках темы рассматриваются возможности python по форматированию строк: %-форматирование, метод format, f-строки. В рамках темы рассматриваются основы работы с регулярными выражениями: базовый синтаксис, примеры. Модуль <i>re</i> в Python. Примеры использования регулярных выражений. В рамках темы рассматривается использование хэширования при работе со строками. Строки в библиотеке <i>pumpru</i>.	Устный опрос. тестирование
6.	Введение в обработку текста на естественном языке в задачах обработки данных	В рамках темы рассматриваются сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python. Использование мемоизации на примере работы со строками. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python. Векторное	Устный опрос. тестирование

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF.	
7.	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.	В рамках темы рассматривается профилирование реализации алгоритмов на Python, принципы решения задачи оптимизации производительности алгоритма. Библиотека Numba: принципы работы, базовые примеры использования. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции	Устный опрос. тестирование
8.	Взаимодействие с базой данных в приложениях обработки данных.	В рамках темы рассматривается взаимодействие из Python с базой данных на примере API SQLite. Базовые возможности работы с транзакциями	Устный опрос. тестирование
9.	Параллельная обработка данных	В рамках темы рассматривается специфика современного аппаратного обеспечения для обработки больших данных и проблема масштабируемости параллельных вычислений. Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение. Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения. Модели параллельного программирования и их сочетаемость с архитектурами параллельных вычислительных систем. Специфика различия между потоками и процессами. Проблема Global Interpreter Lock в Python и способы обхода ее ограничений. Модуль Python multiprocessing – назначение и основные возможности, API multiprocessing.Pool.	Устный опрос. тестирование
10.	Библиотека Dask	В рамках темы рассматривается библиотека для анализа больших объемов данных Python Dask, различные предлагаемые ей подходы к обработке данных. В частности, три ключевых структуры данных Dask: Dask.Array, Dask.DataFrame и Dask.Bag их специфика и принцип выбора структур данных при решении задач. Рассматривается граф зависимостей задач, как ключевая структура для организации параллельной	Устный опрос. тестирование

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		обработки данных в Python Dask. Рассматривается принцип и примеры использования распараллеливание алгоритмов с помощью dask.delayed . Рассматривается структура данных Dask.Array, специфика ее реализации и применения, процедура создания, поддерживаемые Dask.Array операции и ее отличия от NumPy ndarray. Рассматривается структура данных Dask.DataFrame, специфика ее реализации и применения, процедура создания, ограничения использования Dask.DataFrame. Рассматриваются операции мэппинга в Dask.DataFrame и операции Dask.DataFrame работающие со скользящим окном. Рассматривается структура данных Dask.Bag, специфика ее реализации и применения, процедура создания, поддерживаемые Dask.Bag операции. Организация вычислений с помощью Map / Filter / Reduce : общий принцип и специфика параллельной реализации обработки данных с помощью Dask.Bag.	
11.	Обзор проблем обработки больших данных и вычисления общего назначения на GPU	Большие данные – определение и причины возникновения задач обработки больших данных. Вызовы «Больших данных»: объем данных, слабая структурированность данных, связность данных, обработка данных с помощью независимых сервисов. Специфика аппаратного обеспечения для решения задач обработки больших данных. Проблема выбора типичных средств обработки данных, адекватных различным объемам данных. Принцип обработки данных на базе операций map / filter / reduce, принципы архитектуры hadoop. Источники больших данных и прикладные задачи обработки больших данных. История развития и общая характеристика GPU. Архитектура Nvidia CUDA. Принципы организации вычислений в архитектуре Nvidia CUDA.	Устный опрос. тестирование

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		Знакомство с библиотекой PyTorch. Понятие тензора в PyTorch. Базовые операции с тензорами в PyTorch.	

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.2 Занятия семинарского типа

Занятия семинарского типа не предусмотрены

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.3 Лабораторные занятия

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Библиотека NumPy и Pandas	– Технологический стек Python для обработки и анализа данных – Возможности Python как glue language – Организация массивов в NumPy: хранение данных, создание массивов – Принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. – Организация Pandas DataFrame и организация индексации для DataFrame и Series. – Применение универсальных функций и работа с пустыми значениями в Pandas. – Объединение данных из нескольких Pandas DataFrame: общая логика и примеры.	Интерактивная форма, работа на компьютере
2.	Использование различных форматов файлов в задачах обработки данных.	– Формат файлов Pickle, представление данных в этом формате и взаимодействие с ним в Python. – Формат файлов JSON, представление данных в этом формате и взаимодействие с ним в Python. – Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM – Работа с XML с помощью библиотеки BeautifulSoup.	Интерактивная форма, работа на компьютере Интерактивная форма, работа на компьютере

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
3.	Взаимодействие с табличными данными в приложениях обработки данных.	– Взаимодействие с Excel из Python с помощью библиотеки XLWings. – Формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python	Интерактивная форма, работа на компьютере
4.	Визуализация данных	– Построение визуализаций с помощью библиотеки matplotlib – Построение визуализаций с помощью библиотеки pandas – Построение визуализаций с помощью библиотеки seaborn	Интерактивная форма, работа на компьютере
5.	Работа со строками в приложениях обработки данных	– Основы работы с регулярными выражениями: базовый синтаксис, примеры. Модуль re в Python.	Интерактивная форма, работа на компьютере
6.	Введение в обработку текста на естественном языке в задачах обработки данных	– Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python. – Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python.	Интерактивная форма, работа на компьютере
7.	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.	– профилирование реализации алгоритмов на Python – принципы решения задачи оптимизации производительности алгоритма – Библиотека Numba: принципы работы, базовые примеры использования.	Интерактивная форма, работа на компьютере
8.	Взаимодействие с базой данных в приложениях обработки данных.	– Взаимодействие из Python с базой данных с помощью API SQLite.	Интерактивная форма, работа на компьютере
9.	Параллельная обработка данных	– специфика современного аппаратного обеспечения для обработки больших данных и проблема масштабируемости параллельных вычислений. – Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения. – Проблема Global Interpreter Lock в Python и способы обхода ее ограничений.	Интерактивная форма, работа на компьютере

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
		– Модуль Python multiprocessing – назначение и основные возможности, API multiprocessing.Pool.	
10.	Библиотека Dask	– Подход к обработке данных с помощью библиотеки Dask. – Структура данных Dask.Array – принцип работы, API, примеры использования. – Структура данных Dask.DataFrame – принцип работы, API, примеры использования. – Структура данных Dask.Bag – принцип работы, API, примеры использования.	Интерактивная форма, работа на компьютере
11.	Обзор проблем обработки больших данных и вычисления общего назначения на GPU	– Вызовы «Больших данных»: объем данных, слабая структурированность данных, связность данных, обработка данных с помощью независимых сервисов. – Источники больших данных и прикладные задачи обработки больших данных. – Архитектура Nvidia CUDA. Принципы организации вычислений в архитектуре Nvidia CUDA.	Интерактивная форма, работа на компьютере

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.4 Примерная тематика курсовых работ (проектов)

Курсовые работы не предусмотрены.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Изучение теоретического материала	Методические указания по организации самостоятельной работы студентов, утвержденные кафедрой информационных технологий, протокол №1 от 30.08.2019
2	Решение задач	Методические указания по организации самостоятельной работы студентов, утвержденные кафедрой информационных технологий, протокол №1 от 30.08.2019

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,

– в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа,

– в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

– Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

– Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

– Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

– Информационно-коммуникационные технологии (ИКТ) – расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

– Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

– Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

– Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

– Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

– Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

– Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

– Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

- работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;
- проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;
- анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;
- развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4. Оценочные и методические материалы

4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «название дисциплины».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий, доклада-презентации по проблемным вопросам, разноуровневых заданий, ролевой игры, ситуационных задач (указать иное) и **промежуточной аттестации** в форме вопросов и заданий (указать иное) к экзамену (дифференцированному зачету, зачету).

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Перечень вопросов, заданий, тем для подготовки к текущему контролю

Проведение аудиторной самостоятельной работы предполагает командную работу при подготовке сообщений по анализу литературных источников (книг, статей, материалов конференций) на заданную тему, подготовку ответов на контрольные вопросы по темам дисциплины, подготовку контрольных работ.

Примерные вопросы к контрольной работе

1. Большие данные – определение и причины возникновения задач обработки больших данных
2. Специфика современного аппаратного обеспечения для обработки больших данных и проблема масштабируемости параллельных вычислений
3. Выбор типичных средств обработки данных, адекватных различным объемам данных; принцип обработки данных на базе операций map / filter / reduce

4. Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение
5. Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения
6. Модели параллельного программирования и их сочетаемость с архитектурами параллельных вычислительных систем
7. Различия между потоками и процессами, различие между различными планировщиками в Dask
8. Граф зависимостей задач – суть структуры данных, ее построение и использование в Dask
9. Три ключевых структуры данных Dask: их специфика и принцип выбора структуры данных при решении задач
10. Dask.Array – структура данных, специфика реализации и применения, процедура создания
11. Dask.Array – поддерживаемые операции и отличия от NumPy ndarray
12. Распаралеливание алгоритмов с помощью `dask.delayed` – принцип и примеры использования
13. Дополнительные параметры декоратора `dask.delayed` – назначение и примеры использования
14. Использование `dask.delayed` для объектов и операции над объектами `dask.delayed`, включая ограничения их использования
15. Dask.DataFrame – структура данных, специфика реализации и применения, процедура создания Dask.DataFrame
16. Ограничения использования Dask.DataFrame и операции мэппинга в Dask.DataFrame
17. Поддержка Dask.DataFrame операций, работающих со скользящим окном
18. Совместное использование промежуточных результатов в Dask: принцип работы и примеры использования
19. Dask.Bag – структура данных, специфика реализации и применения, процедура создания Dask.Bag
20. Организация вычислений с помощью Map / Filter / Reduce : общий принцип и специфика параллельной реализации обработки данных в Dask.Bag
21. API Dask.Bag – функции мэппинга, фильтрации и преобразования

Примерные задания контрольной работы

Задание 1

1. В массиве чисел, хранящихся в файле `finance.hdf5`, найти строку (вывести ее индекс и содержащиеся значения), в которой более всего значений, превышающих среднее значение по всему массиву. Для расчётов использовать `dask.array`.
2. В массиве чисел, хранящихся в файле `finance.hdf5`, подсчитать количество строк, в которых более 600 значений больше среднего значения по всему массиву. Для расчётов использовать `dask.array`.
3. В массиве чисел, хранящихся в файле `finance.hdf5`, подсчитать количество значений, не отклоняющихся от среднего значения более чем на 3 стандартных отклонения. Для расчетов использовать `dask.array`

Задание 2

1. В accounts.*.csv найти id, для которого в столбце amount встречается наибольшее количество значений, кратных трем. Выполнить задание с использованием Dask, распараллелив процесс обработки данных
2. В accounts.*.csv найти id, для которого сумма положительных значений в столбце amount наибольшая. Выполнить задание с использованием Dask, распараллелив процесс обработки данных.
3. В accounts.*.csv найти id, для которого в столбце amount встречается наибольшее количество значений между 1000 и 1500. Выполнить задание с использованием Dask, распараллелив процесс обработки данных.

Задание 3

Датасет: all_k.zip

Подсчитать, сколько раз в текстовых файлах, лежащих в all_k.zip, встречаются предложения трех видов: вопросительные (в окончании имеют вопросительный знак), побудительные (в окончании имеют восклицательный знак и не имеют вопросительного) и повествовательные (в окончании имеют точку или троеточие, при этом нужно исключить учет точек, встречающихся в сокращениях, таких как "т.к.").

Выполнить задание с использованием Dask (корректным!), распараллелив процесс обработки данных (использование Dask должно приводить к истинной параллельной обработке данных).

Задание 4

Датасет: all_k.zip

Подсчитать, сколько раз встречается каждое из личных местоимений в именительном падеже (полный список: я, ты, он, она, оно, мы, вы, они) в текстовых файлах, лежащих в папке: all_k.zip.

Выполнить задание с корректным использованием Dask, распараллелив процесс обработки данных (использование Dask должно приводить к истинной параллельной обработке данных).

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Примерные вопросы для подготовки к экзамену

1. Большие данные – определение и причины возникновения задач обработки больших данных
2. Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение»

3. Специфика текстовых и бинарных файлов, форматы файлов CSV и Pickle, представление данных в этих форматах и взаимодействие с ними в Python
4. Задача сериализации и десериализации, описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python
5. Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup
6. Форматы файлов NPY и HDF общая характеристика, пример взаимодействия с данными этих форматов в Python
7. Взаимодействие из Python с базой данных на примере API SQLite, базовые возможности работы с транзакциями
8. Взаимодействие с Excel из Python с помощью XLWings: принципы работы и примеры использования
9. Основы работы с регулярными выражениями: базовый синтаксис, примеры использования модуля re в Python
10. Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python
11. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python
12. Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF
13. Модуль multiprocessing – назначение и основные возможности, API multiprocessing.Pool
14. Различия между потоками и процессами, различие между различными планировщиками в Dask
15. Граф зависимостей задач – суть структуры данных, ее построение и использование в Dask
16. Три ключевых структуры данных Dask: их специфика и принцип выбора структуры данных при решении задач
17. Dask.Array – структура данных, специфика реализации и применения, процедура создания
18. Dask.Array – поддерживаемые операции и отличия от NumPy ndarray
19. Распараллеливание алгоритмов с помощью dask.delayed – принцип и примеры использования
20. Дополнительные параметры декоратора dask.delayed – назначение и примеры использования
21. Использование dask.delayed для объектов и операции над объектами dask.delayed, включая ограничения их использования
22. Dask.DataFrame - структура данных, специфика реализации и применения, процедура создания Dask.DataFrame
23. Ограничения использования Dask.DataFrame и операции мэппинга в Dask.DataFrame
24. Поддержка Dask.DataFrame операций, работающих со скользящим окном
25. Совместное использование промежуточных результатов в Dask: принцип работы и примеры использования
26. Dask.Bag - структура данных, специфика реализации и применения, процедура создания Dask.Bag
27. Организация вычислений с помощью Map / Filter / Reduce: общий принцип и специфика параллельной реализации обработки данных в Dask.Bag
28. API Dask.Bag – функции мэппинга, фильтрации и преобразования

29. API Dask.Bag – функции группировки и свертки.

Пример экзаменационного билета

1. (20 баллов) Большие данные – определение и причины возникновения задач обработки больших данных.

2. (20 баллов) Датасет: Chinook_Sqlite.sqlite

С помощью кода на Python с использованием sqlite3 и SQL решить задачу. Реализовать функции на Python:

1. Которая возвращает все имеющиеся плейлисты.

2. Которая по имени плейлиста возвращает количество треков в нем и их суммарную продолжительность.

3. (20 баллов) Датасет: all_k.zip

Подсчитать, сколько раз во всех текстовых файлах, лежащих в all_k.zip, встречаются реплики прямой речи, оформленные в виде диалога (В этом случае каждая реплика начинается с новой строки, перед репликами ставится тире (перед тире возможны различные пробельные символы)). Выполнить задание с использованием Dask, распараллелив процесс обработки данных.

Выполнить задание с использованием Dask (корректным!), распараллелив процесс обработки данных (использование Dask должно приводить к истинной параллельной обработке данных).

4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

5.1 Основная литература:

1. Колдаев, В. Д. Структуры и алгоритмы обработки данных : учебное пособие / В. Д. Колдаев. - Москва : РИОР : ИНФРА-М, 2021. - 296 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 07.12.2023). – Текст : электронный.
2. Сенько, А. Работа с BIGDATA в облаках. Обработка и хранение данных с примерами из Microsoft / А. Сенько. - СПб.: Питер, 2019. - 448 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 07.12.2023). – Текст : электронный.
3. Нархид, Н. Apache Kafka. Поточковая обработка и анализ данных / Н. Нархид. - СПб.: Питер, 2019. - 320 с. ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 07.12.2023). – Текст: электронный.

5.2 Дополнительная литература:

1. Нагаева, И. А. Основы алгоритмизации и программирования: практикум: учебное пособие / И. А. Нагаева, И. А. Кузнецов. – Москва : Берлин : Директ-Медиа, 2021. – 169 с. – ЭБС Университетская библиотека ONLINE. – URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 07.12.2023). – Текст : электронный.
2. Сенько, А. Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure / А. Сенько. - СПб.: Питер, 2019. - 448 с. ЭБС Университетская библиотека ONLINE. – URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 07.12.2023). – Текст : электронный.

5.3. Периодические издания:

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
2. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

5.4. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>

7. Президентская библиотека им. Б.Н. Ельцина <https://www.prilib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
4. Федеральный портал "Российское образование" <http://www.edu.ru/>;
5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.ucheba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--p1ai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru;>
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля)

1. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
2. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
3. Электронно-библиотечная система Znanium <http://www.znaniy.com>
- 1.0.9 [Электронный ресурс]: сайт. – Режим доступа: <https://pypi.python.org/pypi/pylru>
4. Python Data Analysis Library [Электронный ресурс]: сайт. – Режим доступа: <http://pandas.pydata.org/>
5. Python Documentation [Электронный ресурс]: сайт. – Режим доступа: <http://python.org/doc/>
6. Python Standard Library [Электронный ресурс]: сайт. – Режим доступа: <https://docs.python.org/2/library/>
7. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. – Режим доступа: <http://scikit-learn.org>
8. Официальный сайт продукта <https://www.python.org/>
9. Каталог курсов Интернет Университета Информационных Технологий <http://www.intuit.ru/>
10. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>
11. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>
12. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>
13. Dask User Guide <https://docs.dask.org/en/latest/>
14. Dask User Guide <https://docs.dask.org/en/latest/>
15. Matplotlib User Guide // <https://matplotlib.org/stable/users/index.html>
16. Seaborn User Guide // <https://seaborn.pydata.org/tutorial.html>
-

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

7.1 Перечень информационно-коммуникационных технологий

7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

8. Материально-техническое обеспечение по дисциплине (модулю)

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
1.	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения

2.	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением
3.	Практические занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
4.	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
5.	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
6.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.