



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГЕОДЕЗИИ И КАРТОГРАФИИ
(МИИГАиК)

Факультет геоинформатики и информационной безопасности
Кафедра информатики и геоинформационных технологий



**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 5387FD2AF5B17FED816FDE9A1D4973B1
Владелец: Степанченко Алексей Леонидович
Действителен: с 21.07.2025 по 14.10.2026

«УТВЕРЖДАЮ»

Проректор
А. Л. Степанченко
16 февраля 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
«Б1.В.04 ЯЗЫКИ ПРОГРАММИРОВАНИЯ ДЛЯ РАБОТЫ С БОЛЬШИМИ
ДАНЫМИ»**

Уровень высшего образования: магистратура

Направление подготовки: 09.04.03 Прикладная информатика

Направленность(профиль) подготовки: Анализ больших пространственных данных

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Год набора (приема на обучение): 2024

Срок получения образования: 2 года

Объем:
в зачетных единицах: 3 з.е.
в академических часах: 108 ак.ч.

Москва, 2024

Разработчики:

Старший преподаватель кафедры Информатики и геоинформационных технологий Стрельцов А. С.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 916

Согласование и утверждение

№	Подразделение	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	факультет Геоинформатики и информационной безопасности	Председатель ученого совета факультета	Матерухин А. В.	Рассмотрено	14.12.2023, № 7-23

Содержание

1. Цель и задачи освоения дисциплины (модуля)
2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы
3. Место дисциплины в структуре ОП
4. Объем дисциплины и виды учебной работы
5. Содержание дисциплины
 - 5.1. Разделы, темы дисциплины и виды занятий
 - 5.2. Содержание разделов, тем дисциплины и формы текущего контроля
6. Оценочные материалы текущего контроля
7. Оценочные материалы промежуточной аттестации
8. Порядок проведения промежуточной аттестации
9. Материально-техническое и учебно-методическое обеспечение дисциплины
 - 9.1. Перечень основной и дополнительной учебной литературы
 - 9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся
 - 9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине
 - 9.4. Оборудование и технические средства обучения
10. Методические указания по освоению дисциплины (модуля)

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - формирование у обучающихся комплексных компетенций в области применения современных языков программирования и технологий распределенных вычислений для эффективной обработки, анализа и визуализации больших объемов структурированных и пространственных данных в профессиональной деятельности.

Задачи изучения дисциплины:

- Освоение основ программирования на языке Python и специализированных библиотек для работы с большими данными, включая технологии параллельных и распределенных вычислений.;
- Изучение методов и инструментов обработки больших объемов табличных, векторных и растровых пространственных данных с использованием современных фреймворков и облачных платформ.;
- Формирование навыков оптимизации вычислительных процессов, масштабирования решений и интеграции различных технологий для построения эффективных систем обработки больших данных..

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ПК-1 Способен выполнять работы по сбору, обработке и анализу больших пространственных данных

ПК-1.1 Собирает большие пространственные данные

Знать:

ПК-1.1/Зн1 Методы и средства поиска и получения, систематизации и обработки аэрокосмической информации и снимков местности

Уметь:

ПК-1.1/Ум1 Планировать процессы и этапы обработки(получения) данных фотографической, оптико-электронной и других видов аэрокосмических съемок для решения определенного круга производственных и научных задач

Владеть:

ПК-1.1/Нв1 Методами планирования аэросъемочного процесса для получения изображений с заданными параметрами

ПК-1.2 Обработывает большие пространственные данные

Знать:

ПК-1.2/Зн1 Основы и физические принципы регистрации аэрокосмической информации о местности

ПК-1.2/Зн2 Технологии и алгоритмы сбора, обработки и анализа больших данных о Земле

Уметь:

ПК-1.2/Ум1 Осуществлять выбор аэрокосмических снимков, получаемых различными системами и комплексами дистанционного зондирования, в зависимости от решаемых практических или исследовательских задач

ПК-1.2/Ум2 Применять типовые технологии и алгоритмы сбора, обработки и анализа больших данных о Земле для решения практических задач

Владеть:

ПК-1.2/Нв1 Методами сбора и оценки качества пространственной информации, имеющей значения для подготовки и реализации проектов в соответствующих сферах профессиональной деятельности

ПК-1.2/Нв2 Навыками использования типовых технологий и алгоритмов сбора, обработки и анализа больших данных о Земле для решения практических задач

ПК-1.3 Анализирует большие пространственные данные

Знать:

ПК-1.3/Зн1 Основные виды и способы дистанционного зондирования и характеристики формируемых с их помощью снимков и изображений

ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных

ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных

Уметь:

ПК-1.3/Ум1 Ум1 проводить оценку фотографического качества аэрокосмических изображений и анализ соответствия его заданным в ТЗ на съемку параметрам

ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных

ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности

Владеть:

ПК-1.3/Нв1 Навыками оценки качественных характеристик используемой при аэросъемке фотографической аппаратуры и соответствия их решаемой задаче

ПК-1.3/Нв2 Навыками использования типовых алгоритмов анализа больших данных для решения практических задач профессиональной деятельности

ПК-1.3/Нв3 Навыками использования специализированного программного обеспечения для анализа больших данных

3. Место дисциплины в структуре ОП

Дисциплина (модуль) Б1.В.04 «Языки программирования для работы с большими данными» относится к формируемой участниками образовательных отношений части образовательной программы и изучается в семестре(ах): 1.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

4. Объем дисциплины (модуля) и виды учебной работы

Период обучения	Общая трудоемкость (часы)	Общая трудоемкость (ЗЕТ)	Лекционные занятия (часы)	Практические занятия (часы)	Самостоятельная работа (часы)	Промежуточная аттестация (часы)
1 семестр	108	3	34	34	40	Зачет
Всего	108	3	34	34	40	

5. Содержание дисциплины (модуля)

5.1. Разделы, темы дисциплины и виды занятий
(часы промежуточной аттестации не указываются)

Наименование раздела, темы	Всего	Лекционные занятия	Практические занятия	Самостоятельная работа	Планируемые результаты обучения, соотношенные с результатами освоения программы
Раздел 1. Основы Python для обработки больших данных	14	4	4	6	ПК-1.2
Тема 1.1. Введение в язык Python и особенности работы с большими данными	4	1	1	2	
Тема 1.2. Управляющие конструкции и обработка исключений	4	1	1	2	
Тема 1.3. Модули, пакеты и генераторы	6	2	2	2	
Раздел 2. Объектно-ориентированное программирование для масштабируемых решений	8	2	2	4	ПК-1.2
Тема 2.1. ООП-подход к обработке больших данных	4	1	1	2	
Тема 2.2. ООП-рефакторинг и оптимизация	4	1	1	2	
Раздел 3. NumPy и Pandas для работы с большими табличными данными	10	3	3	4	ПК-1.1 ПК-1.2 ПК-1.3
Тема 3.1. NumPy для эффективных численных вычислений	4	1	1	2	
Тема 3.2. Pandas для анализа больших датасетов	6	2	2	2	
Раздел 4. Распределенные вычисления с Dask	18	6	6	6	ПК-1.2
Тема 4.1. Введение в Dask и параллельные вычисления	6	2	2	2	
Тема 4.2. Dask DataFrame для табличных данных	6	2	2	2	
Тема 4.3. Dask Delayed для параллелизации функций	6	2	2	2	
Раздел 5. Обработка больших пространственных векторных данных	18	6	6	6	ПК-1.1 ПК-1.2 ПК-1.3
Тема 5.1. GeoPandas для векторного анализа	6	2	2	2	
Тема 5.2. Dask-GeoPandas для распределенной обработки	6	2	2	2	

Тема 5.3. API и потоковая обработка геоданных	6	2	2	2	
Раздел 6. Обработка больших растровых данных	16	5	5	6	ПК-1.1 ПК-1.2 ПК-1.3
Тема 6.1. RasterIO для работы с большими растрами	6	2	2	2	
Тема 6.2. Xarray и rioхarray для многомерных данных	6	2	2	2	
Тема 6.3. Распределенная обработка растров	4	1	1	2	
Раздел 7. Облачные платформы для больших геоданных	15	5	5	5	ПК-1.1 ПК-1.2 ПК-1.3
Тема 7.1. Google Earth Engine для обработки данных планетарного масштаба	6	2	2	2	
Тема 7.2. Работа с большими коллекциями в GEE	6	2	2	2	
Тема 7.3. Интеграция облачных и локальных вычислений	3	1	1	1	
Раздел 8. Оптимизация и профилирование	9	3	3	3	ПК-1.2
Тема 8.1. Многопоточность и многопроцессорность	3	1	1	1	
Тема 8.2. Асинхронное программирование	3	1	1	1	
Тема 8.3. Профилирование и оптимизация памяти	3	1	1	1	
Итого	108	34	34	40	

5.2. Содержание разделов, тем дисциплин и формы текущего контроля

Раздел 1. Основы Python для обработки больших данных

Тема 1.1. Введение в язык Python и особенности работы с большими данными

Установка и настройка среды разработки для работы с большими данными . Основные типы и структуры данных Python. Функции ввода-вывода для больших файлов.

Тема 1.2. Управляющие конструкции и обработка исключений

Условные операторы и циклы для обработки больших массивов данных . Создание эффективных функций. Обработка исключений при работе с большими файлами. Контекстные менеджеры для управления ресурсами. Работа с текстовыми файлами по частям.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 1.3. Модули, пакеты и генераторы

Организация кода в модули и пакеты . Генераторы и итераторы для обработки больших данных. Модуль itertools для эффективной итерации. Ленивые вычисления.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 2. Объектно-ориентированное программирование для масштабируемых решений

Тема 2.1. ООП-подход к обработке больших данных

Классы и объекты для представления больших датасетов .Инкапсуляция логики обработки данных. Создание классов для потоковой обработки.

Тема 2.2. ООП-рефакторинг и оптимизация

Наследование и полиморфизм для расширяемых систем. Абстрактные классы для унификации интерфейсов обработки данных.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 3. NumPy и Pandas для работы с большими табличными данными

Тема 3.1. NumPy для эффективных численных вычислений

Векторизованные операции в NumPy . Работа с большими многомерными массивами. Оптимизация использования памяти. Chunking стратегии для обработки массивов, не помещающихся в память.

Тема 3.2. Pandas для анализа больших датасетов

Эффективная работа с DataFrame для больших данных . Чтение больших CSV файлов по частям (chunksize). Оптимизация типов данных для экономии памяти. Использование категориальных типов. Параллельная обработка с pandas.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 4. Распределенные вычисления с Dask

Тема 4.1. Введение в Dask и параллельные вычисления

Архитектура Dask и принципы работы. Dask Arrays для больших численных массивов. Планировщики задач: синхронный, многопоточный, распределенный. Визуализация графа вычислений.

Тема 4.2. Dask DataFrame для табличных данных

Dask DataFrame как масштабируемая альтернатива Pandas. Ленивые вычисления и оптимизация запросов. Партиционирование данных. Операции shuffle и их оптимизация. Интеграция с Pandas workflows.

Тема 4.3. Dask Delayed для параллелизации функций

Декоратор @delayed для параллелизации пользовательских функций. Применение Dask Delayed к функциям пакета geometry. Построение графов вычислений. Оптимизация CPU-интенсивных задач.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 5. Обработка больших пространственных векторных данных

Тема 5.1. GeoPandas для векторного анализа

Работа с большими векторными датасетами в GeoPandas . Пространственные индексы для ускорения операций. Оптимизация пространственных соединений. Буферизация и overlay операции для больших данных.

Тема 5.2. Dask-GeoPandas для распределенной обработки

Масштабирование GeoPandas с помощью dask-geopandas. Партиционирование пространственных данных. Распределенные пространственные операции. Обработка данных OpenStreetMap для больших территорий .

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 5.3. API и потоковая обработка геоданных

Работа с API геосервисов для получения больших объемов данных . Пагинация и батчинг запросов. Кэширование результатов. Асинхронные запросы с aiohttp.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 6. Обработка больших растровых данных

Тема 6.1. RasterIO для работы с большими растрами

Чтение больших растров по окнам (windows) . Пирамиды изображений и overview. Виртуальные растры (VRT). Оптимизация операций чтения/записи.

Тема 6.2. Xarray и rioxarray для многомерных данных

Xarray для работы с многомерными научными данными. Интеграция с Dask для чанкованных массивов. Rioxarray для геопространственных растров. Ленивая загрузка и обработка временных серий спутниковых снимков.

Тема 6.3. Распределенная обработка растров

Dask-чанкинг для xarray/gioxarray. Параллельная обработка больших растровых мозаик. Map-reduce операции над растрами. Оптимизация I/O операций.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 7. Облачные платформы для больших геоданных

Тема 7.1. Google Earth Engine для обработки данных планетарного масштаба

Архитектура Google Earth Engine . Серверная обработка петабайтов данных. Каталоги данных и их эффективное использование. Оптимизация запросов и reduceRegion операций.

Тема 7.2. Работа с большими коллекциями в GEE

Фильтрация и обработка ImageCollection . Временной анализ больших серий. Экспорт результатов в Google Cloud Storage. Батч-обработка и оптимизация квот.

Тема 7.3. Интеграция облачных и локальных вычислений

Гибридные подходы: GEE + локальная обработка. Использование Google Colab для масштабирования . Интеграция с BigQuery для табличных данных.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 8. Оптимизация и профилирование

Тема 8.1. Многопоточность и многопроцессорность

Модули threading и multiprocessing в Python. GIL и его обход. Пулы процессов для CPU-bound задач. Concurrent.futures для высокоуровневого параллелизма.

Тема 8.2. Асинхронное программирование

Asyncio для I/O-bound операций. Асинхронная обработка API запросов. Aiofiles для асинхронной работы с файлами. Интеграция с синхронным кодом.

Тема 8.3. Профилирование и оптимизация памяти

Профилирование с cProfile и line_profiler. Анализ использования памяти с memory_profiler. Оптимизация структур данных. Использование Numba и Cython для критических участков.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

6. Порядок проведения промежуточной аттестации

Порядок проведения промежуточной аттестации регламентируется локальными нормативными актами Университета:

- Положением о формах, периодичности, порядке текущего контроля успеваемости и промежуточной аттестации обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.6 «Организация и проведение промежуточной аттестации».
- Положением о порядке осуществления образовательного процесса, проведения текущего контроля успеваемости и промежуточной аттестации обучающихся с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.8 «Проведение промежуточной аттестации с применением ЭО и ДОТ».

Промежуточная аттестация: Зачет, 1 семестр.

Зачет по дисциплине является формой промежуточной аттестации и проводится в форме тестирования.

Студентам будет предложен тест, содержащий 37 вопросов закрытого типа. На выполнение работы отводится 60 минут, в течение которых использование вспомогательных материалов и электронных устройств запрещено.

Оценивание проводится по принципу «зачтено» или «не зачтено» Для успешной сдачи зачета и получения отметки «зачтено» студенту необходимо дать правильные ответы не менее чем на 60% вопросов, что составляет 23 правильных ответа из 37. В случае, если количество правильных ответов составит менее 23, зачет считается не сданным и выставляется отметка «не зачтено»

7. Оценочные материалы текущего контроля

Раздел 1. Основы Python для обработки больших данных

Контролируемые ИДК: ПК-1.2

Тема 1.2. Управляющие конструкции и обработка исключений

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 1.3. Модули, пакеты и генераторы

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 2. Объектно-ориентированное программирование для масштабируемых решений

Контролируемые ИДК: ПК-1.2

Тема 2.2. ООП-рефакторинг и оптимизация

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 3. NumPy и Pandas для работы с большими табличными данными

Контролируемые ИДК: ПК-1.1 ПК-1.2 ПК-1.3

Тема 3.2. Pandas для анализа больших датасетов

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 4. Распределенные вычисления с Dask

Контролируемые ИДК: ПК-1.2

Тема 4.3. Dask Delayed для параллелизации функций

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 5. Обработка больших пространственных векторных данных

Контролируемые ИДК: ПК-1.1 ПК-1.2 ПК-1.3

Тема 5.2. Dask-GeoPandas для распределенной обработки

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 5.3. API и потоковая обработка геоданных

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 6. Обработка больших растровых данных

Контролируемые ИДК: ПК-1.1 ПК-1.2 ПК-1.3

Тема 6.3. Распределенная обработка растров

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 7. Облачные платформы для больших геоданных

Контролируемые ИДК: ПК-1.1 ПК-1.2 ПК-1.3

Тема 7.3. Интеграция облачных и локальных вычислений

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 8. Оптимизация и профилирование

Контролируемые ИДК: ПК-1.2

Тема 8.3. Профилирование и оптимизация памяти

Форма контроля/оценочное средство: Просмотр выполненных работ

8. Оценочные материалы промежуточной аттестации

1 семестр, Зачет

Контролируемые ИДК: ПК-1.1 ПК-1.2 ПК-1.3

Вопросы/Задания:

1. Какое основное преимущество использования генераторов при обработке больших файлов?

- А. Они работают быстрее обычных функций
- В. Они позволяют выполнять "ленивые" вычисления, обрабатывая данные по частям и экономя память
- С. Они автоматически распараллеливают вычисления
- Д. Они упрощают синтаксис циклов

2. Для чего в Python используется контекстный менеджер (конструкция `with`)?

- А. Для обработки математических исключений
- В. Для гарантированного управления ресурсами, например, автоматического закрытия файлов
- С. Для создания классов и объектов

D. Для импорта модулей и пакетов

3. Какой модуль стандартной библиотеки Python предоставляет функции для создания эффективных итераторов?

- A. collections
- B. functools
- C. itertools
- D. sys

4. Какой подход из перечисленных НЕ является способом обхода Global Interpreter Lock (GIL) для CPU-bound задач?

- A. Использование модуля ``multiprocessing``
- B. Использование модуля ``threading``
- C. Использование Cython
- D. Использование Numba

5. Какой инструмент используется для по-строчного профилирования кода с целью выявления узких мест?

- A. cProfile
- B. memory_profiler
- C. line_profiler
- D. timeit

6. Какой принцип ООП используется для сокрытия сложной логики обработки данных внутри класса?

- A. Наследование
- B. Полиморфизм
- C. Инкапсуляция
- D. Абстракция

7. Какова основная цель использования абстрактных классов при проектировании систем обработки данных?

- A. Ускорить выполнение кода
- B. Уменьшить потребление памяти
- C. Создать единый, унифицированный интерфейс для различных классов-обработчиков
- D. Обеспечить хранение данных

8. Что такое векторизация в NumPy?

- A. Процесс преобразования изображений в векторный формат
- B. Выполнение операций над целыми массивами данных без явных циклов в Python
- C. Создание векторов в трехмерном пространстве
- D. Разделение данных на обучающую и тестовую выборки

9. Для чего предназначен параметр ``chunksize`` при чтении больших CSV-файлов с помощью Pandas?

- A. Для указания размера сжатия файла
- B. Для чтения и обработки файла по частям (чанкам), а не целиком
- C. Для установки количества строк в итоговом DataFrame
- D. Для определения размера буфера в памяти

10. Какой тип данных в Pandas рекомендуется использовать для столбцов с небольшим количеством уникальных строковых значений для экономии памяти?

- A. object
- B. string
- C. category
- D. int64

11. На каком принципе основана работа библиотеки Dask?

- A. На немедленном выполнении всех вычислений
- B. На построении графа задач и "ленивых" вычислениях
- C. На использовании SQL-подобного синтаксиса
- D. На компиляции Python-кода в C++

12. Какой декоратор Dask используется для распараллеливания выполнения обычных Python-функций?

- A. @dask.compute
- B. @dask.parallel
- C. @dask.delayed
- D. @dask.visualize

13. Чем Dask DataFrame принципиально отличается от Pandas DataFrame?

- A. Dask DataFrame хранится только на диске
- B. Dask DataFrame состоит из множества Pandas DataFrame (партиций) и выполняет операции лениво
- C. Dask DataFrame имеет больше методов для визуализации
- D. Dask DataFrame не поддерживает индексацию

14. Какой планировщик задач Dask следует использовать для выполнения вычислений на кластере из нескольких машин?

- A. Синхронный (synchronous)
- B. Многопоточный (threaded)
- C. Распределенный (distributed)
- D. Многопроцессорный (multiprocessing)

15. Какова основная цель пространственного индекса в GeoPandas?

- A. Ускорение пространственных запросов и соединений
- B. Изменение системы координат
- C. Визуализация геоданных на карте
- D. Уменьшение размера файла

16. Для какой задачи предназначена библиотека 'dask-geopandas'?

- A. Для создания интерактивных карт
- B. Для масштабирования операций GeoPandas на наборы данных, превышающие объем оперативной памяти
- C. Для преобразования растровых данных в векторные
- D. Для работы с 3D-данными

17. Что такое "пагинация" при работе с API геосервисов?

- A. Процесс кэширования результатов запросов
- B. Процесс получения данных в виде отдельных "страниц" или порций
- C. Процесс асинхронного выполнения запросов
- D. Процесс преобразования форматов данных

18. Какая библиотека используется для чтения больших растровых файлов по частям (окнам)?

- A. GeoPandas
- B. RasterIO
- C. Xarray
- D. Ruproj

19. Что такое виртуальный растр (VRT)?

- A. Растр, полностью загруженный в оперативную память
- B. Файл, описывающий мозаику или трансформации над другими растрами без создания нового файла с пикселями
- C. Растр, сжатый с потерями
- D. Временный файл, создаваемый при обработке

20. Какая библиотека предназначена для работы с многомерными размеченными массивами данных и хорошо интегрируется с Dask?

- A. NumPy
- B. Pandas
- C. Xarray
- D. Matplotlib

21. В чем заключается основная особенность архитектуры Google Earth Engine (GEE)?

- A. Вся обработка данных происходит на локальном компьютере пользователя
- B. Обработка петабайтов геоданных происходит на серверах Google
- C. GEE предназначен только для визуализации данных
- D. GEE работает только с векторными данными

22. Что делает функция `reduceRegion` в Google Earth Engine?

- A. Уменьшает разрешение изображения
- B. Обрезает изображение по заданной геометрии
- C. Агрегирует значения пикселей в пределах указанного региона (например, вычисляет среднее)
- D. Экспортирует результат в Google Drive

23. Что представляет собой `ImageCollection` в Google Earth Engine?

- A. Один спутниковый снимок с несколькими каналами
- B. Коллекцию векторных геометрий
- C. Стек или временную серию изображений
- D. Результат классификации изображения

24. Для какого типа задач наиболее эффективно применение `asyncio`?

- A. Для задач, интенсивно использующих CPU (CPU-bound)
- B. Для задач, связанных с ожиданием операций ввода-вывода (I/O-bound), например, сетевых запросов
- C. Для любых задач, требующих распараллеливания
- D. Для задач, требующих большого объема оперативной памяти

25. Какая технология используется для ускорения Python-кода путем его JIT-компиляции в машинный код?

- A. Cython
- B. Numba
- C. Multiprocessing
- D. PyPy

26. Какой формат хранения данных является колоночным и хорошо подходит для аналитических запросов к большим датасетам?

- A. CSV
- B. JSON
- C. Parquet
- D. Feather

27. Что такое "ленивые вычисления"?

- A. Вычисления, которые выполняются очень медленно
- B. Откладывание вычислений до тех пор, пока их результат действительно не понадобится
- C. Вычисления, выполняемые в фоновом потоке
- D. Вычисления с пониженным приоритетом

28. Какая библиотека является основной для работы с геопространственными векторными данными в экосистеме Python?

- A. RasterIO
- B. GeoPandas
- C. Shapely
- D. Fiona

29. Какой инструмент используется для анализа потребления памяти Python-скриптом?

- A. cProfile
- B. line_profiler
- C. memory_profiler
- D. time

30. Какой формат данных разработан для быстрой записи и чтения DataFrame между R и Python?

- A. Parquet
- B. HDF5

- C. Zarr
- D. Feather

31. Какая библиотека позволяет использовать ``async/await`` для асинхронной работы с файлами?

- A. os
- B. aiohttp
- C. aiofiles
- D. threading

32. Что такое "чанкинг" (chunking) применительно к массивам NumPy или Xarray?

- A. Сортировка массива
- B. Разделение большого массива на небольшие блоки (чанки) для обработки по частям
- C. Изменение типа данных массива
- D. Транспонирование массива

33. Какая библиотека является зависимостью ``rioxarray`` и обеспечивает ему геопространственные функции?

- A. GeoPandas
- B. RasterIO
- C. Dask
- D. Matplotlib

34. Какой гибридный подход к вычислениям упоминается в программе курса?

- A. Использование Spark и Dask одновременно
- B. Совмещение облачной обработки в GEE и локальной обработки на Python
- C. Использование CPU и GPU для одной задачи
- D. Написание кода на Python и Java в одном проекте

35. Что такое GIL (Global Interpreter Lock)?

- A. Механизм, который позволяет Python выполнять несколько потоков на разных ядрах CPU одновременно
- B. Механизм в CPython, который не позволяет нескольким нативным потокам исполнять байт-код Python одновременно
- C. Библиотека для шифрования данных
- D. Инструмент для отладки кода

36. Какой формат данных предназначен для хранения чанкованных, сжатых N-мерных массивов и является облачно-оптимизированной альтернативой HDF5?

- A. Parquet
- B. Zarr
- C. Feather
- D. CSV

37. Какая библиотека используется для асинхронных HTTP-запросов?

- A. requests
- B. urllib3
- C. aiohttp
- D. socket

9. Материально-техническое и учебно-методическое обеспечение дисциплины

9.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Железнов М. М. Методы и технологии обработки больших данных: учебно-методическое пособие: учебно-методическое пособие / М. М. Железнов. - Методы и технологии обработки больших данных - Москва: МИСИ-МГСУ, ЭБС АСВ, 2020. - 46 с. - 978-5-7264-2193-3. - Текст: электронный. // IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/101802.html> (дата обращения: 16.03.2025). - Режим доступа: по подписке

2. Воронов В. И. Data Mining - технологии обработки больших данных: учебное пособие: учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. - Data Mining - технологии обработки больших данных - Москва: Московский технический университет связи и информатики, 2018. - 47 с. - 2227-8397. - Текст: электронный. // IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/81324.html> (дата обращения: 16.03.2025). - Режим доступа: по подписке

Дополнительная литература

1. Макшанов А. В. Большие данные. Big Data: учебник для вузов: учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. - 4-е изд., стер. - Санкт-Петербург: Лань, 2024. - 188 с. - 978-5-507-47346-5. - Текст: электронный. // Издательство Лань: [сайт]. - URL: <https://e.lanbook.com/img/cover/book/362318.jpg> (дата обращения: 16.03.2025). - Режим доступа: по подписке

9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных

Не используются.

Ресурсы «Интернет»

1. <https://docs.python.org/3/index.html> - Техническая документация по языку программирования Python
2. <https://sdo.miigaik.ru/> - Система дистанционного обучения МИИГАиК

9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения

(обновление производится по мере появления новых версий программы)

1. Anaconda;
2. Microsoft Visual Studio;

Перечень информационно-справочных систем

(обновление выполняется еженедельно)

1. 1С Предприятие;

9.4. Оборудование и технические средства обучения

Учебная аудитория

- Маркерная доска - 1 шт.
- Моноблок - 96 шт.
- Проектор - 8 шт.
- Автоматизированное рабочее место (АРМ) - 40 шт.
- Плоттер - 1 шт.
- Сканер - 1 шт.

Интерактивная панель - 1 шт.

10. Методические указания по освоению дисциплины (модуля)

С целью успешного освоения данного предмета студенту необходимо выполнять домашние работы по указанным преподавателям темам и представленным на занятиях материалам. Для более успешного выполнения работы студенту предлагается использование учебно-методических материалов, соответствующих указанной теме.

Самостоятельная работа студентов описывается и регулируется методическими рекомендациями по дисциплине.

Описание возможностей изучения дисциплины лицами с ОВЗ и инвалидами

При необходимости обучения студентов-инвалидов и лиц с ограниченными возможностями здоровья аудиторные занятия могут быть заменены или дополнены изучением полнотекстовых лекций, презентаций, видео- и аудиоматериалов в электронной информационно-образовательной среде (ЭИОС) университета. Индивидуальные задания подбираются в адаптированных к ограничениям здоровья формах (письменно или устно, в форме презентаций). Выбор методов обучения зависит от их доступности для инвалидов и лиц с ограниченными возможностями здоровья.

Рабочая программа дисциплины для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.