1. Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение
2. Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения
3. Проблема Global Interpreter Lock в Python и способы обхода ее ограничений
4. Модели параллельного программирования и их сочетаемость с архитектурами параллельных вычислительных систем
5. Профилирование реализации алгоритмов на Python, принципы решения задачи оптимизации производительности алгоритма
6. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции
7. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
8. Технологический стек Python для обработки и анализа данных, Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python
9. Универсальные функции и применение функций по осям в NumPy
10. Маскирование и прихотливое индексирование в NumPy
11. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
12. Numba: принципы работы, базовые примеры использования
13. Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение»
14. Применение универсальных функций и работа с пустыми значениями в Pandas
15. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры
16. Организация Pandas DataFrame и организация индексации для DataFrame и Series
17. Специфика текстовых и бинарных файлов, форматы файлов CSV и Pickle, представление данных в этих форматах и взаимодействие с ними в Python
18. Задача сериализации и десериализации, описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python
19. Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup
20. Форматы файлов NPY и HDF общая характеристика, пример взаимодействие с данными этих форматов в Python
21. Взаимодействие с Excel из Python с помощью XLWings: принципы работы и примеры использования
22. Основы работы с регулярными выражениями: базовый синтаксис, примеры использования модуля re в Python
23. Модуль multiprocessing – назначение и основные возможности, API multiprocessing.Pool
24. Сегментация и токенезация текста на естественном языке, стеммминг и лемматизация, примеры на Python
25. Расстояние Левеншнтейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python
26. Различия между потоками и процессами, различие между различными планировщиками в Dask
27. Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF
28. Граф зависимостей задач – суть структуры данных, ее построение и использование в Dask
29. Dask.Array – структура данных, специфика реализации и применения, процедура создания
30. Dask.Array – поддерживаемые операции и отличия от NumPy ndarray
31. Dask.Bag - структура данных, специфика реализации и применения, процедура создания DaskBag
32. API Dask.Bag – функции группировки и свертки
33. рганизация вычислений с помощью Map / Filter / Reduce : общий принцип и специфика параллельной реализации обработки данных в Dask.Bag
34. API Dask.Bag – функции мэппинга, фильтрации и преобразования