

ЛАБОРАТОРНАЯ РАБОТА №4

Курс «Параллельные методы и алгоритмы»



Тема: Распределенные вычисления MapReduce.

Цель: Научиться писать и отлаживать распределенные приложения на основе методологии MapReduce.

Темы для предварительной проработки ^[устно]:

- Методология MapReduce.
- Паттерны MapReduce.
- Фреймворки Hadoop / Spark / Hive.
- Библиотека mrjob.

Индивидуальные задания ^[код] :

- 1) Написать MapReduce-приложение, решающее задачу №1 из лабораторной работы №1.
- 2) Написать MapReduce-приложение для подсчета количества уникальных слов в тексте.
- 3) Написать MapReduce-приложение для выдачи N наиболее встречаемых слов в тексте.
- 4) Написать MapReduce-приложение для выдачи N наиболее встречаемых пар слов в тексте.
- 5) Написать MapReduce-приложение для выдачи N наиболее встречаемых слов в тексте, сгруппированных по длине слова (т.е. N 3-буквенных слов, N 4-буквенных и т.д.).

Контрольные вопросы ^[отчет] :

1. В чем разница между вертикальным и горизонтальным масштабированием системы?
2. Раскройте суть подхода MapReduce.
3. Приведите примеры задач, где не нужна процедура Reduce.
4. Кратко опишите основные паттерны MapReduce.
5. Перечислите основные компоненты фреймворка Hadoop.
6. Что представляет собой файловая система HDFS?
7. Как работать со streaming-интерфейсом Hadoop?
8. Кратко опишите назначение фреймворка Spark.
9. Кратко опишите назначение фреймворка Hive
10. Как реализует подход MapReduce библиотека mrjob?

Рекомендуемые источники:

- [1] Гергель В.П. Теория и практика параллельных вычислений: учеб. пособие / В.П. Гергель. – М.: Интернет-Ун-т информ. технологий: БИНОМ. Лаб. знаний, 2007. – 423 с.
- [2] Data Intensive Text Processing with MapReduce [Электронный ресурс] / URL: <http://lintool.github.io/MapReduceAlgorithms/MapReduce-book-final.pdf>.
- [3] Herlihy M. The Art of Multiprocessor Programming / M. Herlihy, N. Shavit. – Morgan Kaufmann, 2012. – 536 p.
- [4] Воеводин, В. В. Параллельные вычисления : Учеб. пособие для вузов по направлению "Прикладная математика и информатика" / В. В. Воеводин, Вл. В. Воеводин. – СПб. : БХВ-Петербург, 2002. – 608 с.