

Курсовой проект  
*Фреймворк и файловая система для  
распределённой обработки больших данных в  
рамках концепции map-reduce*

Выборнов А.И,

МГТУ им. Н. Э. Баумана

art-vybor@ya.com

17 октября 2014 г.

# Обзор

Постановка задачи

Концепция map-reduce

Архитектура

Отчёт

# Постановка задачи

- ▶ Анализ требований и проектирование архитектуры системы. Реализация нераспределённого map-reduce. Решение проблемы RPC.
- ▶ Реализация распределённой файловой системы. Реализация фреймворка. Тестирование на примерах.

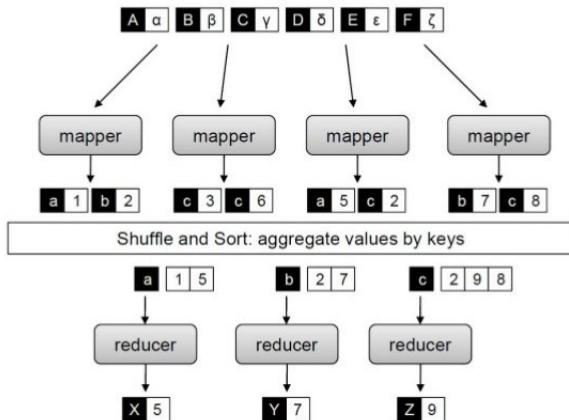
# Зачем нужен map-reduce?

- ▶ Обработка больших данных (Big Data).
  - ▶ Вычисления превосходят возможности одной машины.
  - ▶ Данные не помещаются в памяти, необходимо обращаться к диску.
  - ▶ Можно хранить много данных, но задержки и пропускная способность оборудования растут пропорционально данным.
- ▶ Удобная абстракция для построения алгоритмов обработки больших данных.
- ▶ Устойчивость к отказам.

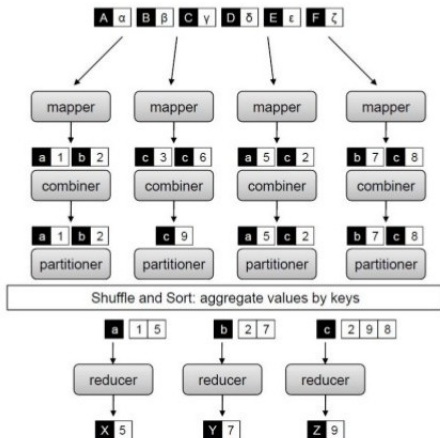
# Что такое map-reduce?

- ▶ Структура  $(key, value)$  - пара (ключ, значение).
- ▶ Программирование представляет собой определение двух функций:
  - ▶  $map : (key, value) \rightarrow [(key, value)]$
  - ▶  $reduce : (key, [value]) \rightarrow [(key, value)]$
- ▶ Между стадиями *map* и *reduce* происходит группировка и сортировка данных.

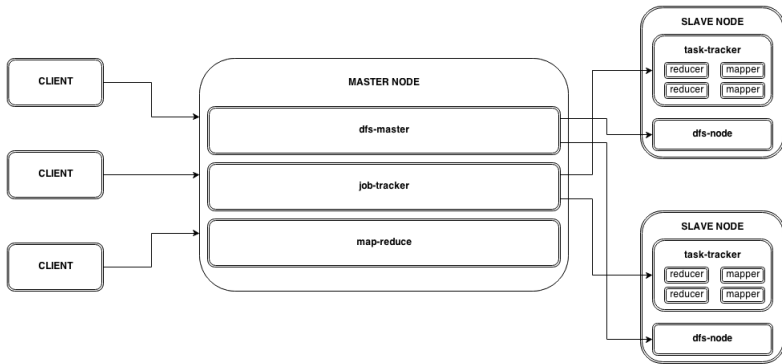
# Что такое map-reduce?



# А на самом деле

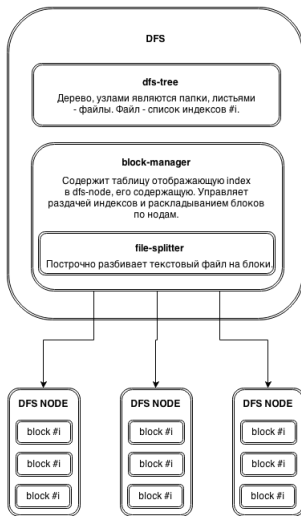


# Архитектура распределённого map-reduce





# Архитектура распределённой файловой системы



## Что сделано

- ▶ Разработаны архитектуры распределённых map-reduce и файловой системы.
- ▶ Реализован нераспределённый map-reduce, включающий в себя все этапы распределённого.
- ▶ Для реализации RPC была выбрана связка: ZeroMQ + Pickle.
- ▶ Реализована распределённая файловая система.

# Что предстоит сделать

- ▶ Развить существующую распределённую файловую систему.
- ▶ Реализовать распределённый map-reduce.
- ▶ Протестировать полученную систему на больших данных и проанализировать результаты.

Курсовой проект  
*Фреймворк и файловая система для  
распределённой обработки больших данных в  
рамках концепции map-reduce*

Выборнов А.И,

МГТУ им. Н. Э. Баумана

art-vybor@ya.com

17 октября 2014 г.