

Домашнее задание №2

Срок сдачи: 4 ноября, 23:59:59. Решения, присланные позже данного срока, не принимаются.

Обратите внимание, что в данном задании оценивается временная эффективность решения!

Формулировка задания: На вход подается .csv файл (разделители – запятые) с финансовыми данными, доступный по адресу:

wasb://financedata@bigdatamsu.blob.core.windows.net/

Столбцы данного файла имеют следующие названия:

#SYMBOL,SYSTEM,MOMENT,ID_DEAL,PRICE_DEAL,VOLUME,OPEN_POS,DIRECTION

где #SYMBOL – название финансового инструмента;

MOMENT – время (дата);

PRICE_DEAL – цена.

Пример строки в файле:

SVH1,F,20110111100000080,255223067,30.46000,1,8714,S

Данный файл отсортирован по дате и времени.

Внимание! Так как данный файл с финансовыми данными очень большого размера, отладку программы следует производить на данных меньшего размера и только убедившись, что все работает, запускать программу на данных большого размера.

Указанный файл с финансовыми данными меньшего размера Вы можете скачать по адресу http://bigdata.cs.msu.ru/images/9/99/Finance_example.zip

Вам необходимо:

- 1) Найти максимальную и минимальную цену каждого финансового инструмента за каждый день
- 2) Упорядочить найденную в первом пункте статистику по дате (и времени) и инструменту. Данную подзадачу можно решать как без использования hadoop (используя `gnu sort`), так и с hadoop (при решении с hadoop Вам будут начислены дополнительные бонусные баллы)
- 3) Сделать отчет о переданном по сети трафике и времени исполнения для двух и четырех рабочих АЗ-узлов в кластере, подобрать оптимальное число редьюсеров для обоих вариантов по времени исполнения. Включить в отчет таблицы по

результатам проведенных экспериментов, сформулировать соответствующие выводы.

Далее Вам необходимо на почту курса bigdata@cs.msu.ru отправить архив в формате Task2-Фамилия.jar (фамилия на англ.), содержащий следующие файлы:

- 1) Файл `finance.java`, в котором производится поиск максимальной и минимальной цен каждого финансового инструмента за каждый день, а также сортировка статистики по дате и инструменту (если сделали)
- 2) Сформированный `finance.jar` файл
- 3) Файл `finance.txt` с отсортированной статистикой в формате:
Дата(и время) ПРОБЕЛ инструмент ПРОБЕЛ максимальная_цена ПРОБЕЛ минимальная_цена
- 4) Файл `finance.pdf` с выполненным отчетом
- 5) Вспомогательные файлы для сборки (если используются)
- 6) Файл `readme.txt` с описанием того, как Вы компилировали и запускали программы