Домашнее задание по MapReduce

**DEADLINE: 4.12.2016, 23:59**

При решении задач можно пользоваться примерами, кот. были разобраны на семинарах. Коды доступны по адресу: /home/velkerr/seminars/practice\*. Пример по Hadoop streaming (streaming\_wordcount) будет разобран на ближайшем семинаре.

**Уровень II** (1,5 балла).

Посчитать число слов для каждого персонажа, число реплик и число слов на реплику. Одна реплика персонажа - это несколько подряд идущих фраз (строк) этого персонажа. Погрешностью, возникающей в тех случаях, когда часть реплики попадает в 1 блок, а часть в другой (а значит, и в другой маппер), можно пренебречь. Слова очистить от знаков пунктуации.

**Данные:** текст «Горе от ума» /data/griboedov: персонаж -> строка реплики

**Формат результата:** персонаж, число слов, число реплик, число слов на реплику

**Правила оформления поста при сдаче этого ДЗ**

В посте нужно указать:

1. Аккаунт на кластере
2. Уровень и № задачи.
3. Директорию с кодом на сервере.
4. Как собирать проект
5. Как запускать (семантика аргументов + пример запуска).
6. Директорию в HDFS, где лежит результат работы программы.