

Все программы хранятся в папке Java.

Для компиляции требуется вести следующие команды

- 1) `export HADOOP_CLASSPATH=$(hadoop classpath)`
- 2) `javac -classpath $HADOOP_CLASSPATH -d jar_classes java/*`
- 3) `jar -cvf myMapReduce-1.0-SNAPSHOT.jar -C jar_classes/ .`

В итоге получаем `myMapReduce-1.0-SNAPSHOT.jar`, которые нужно перетащить на отдаленный сервер с помощью `scp`.

Предварительно данные были загруженные из `stackoverflow` в папку `/user/stud/stackoverflow/landing/Posts`.

Запросы будут аналогичны как в Hive.

Запрос 1:

Выбрать те строки таблицы поста, которые являются вопросами (`PostTypeId==1`) и просмотры не меньше 50 тыс.

Это реализовано в `FilterPost.class`.

```
hadoop jar myMapReduce-1.0-SNAPSHOT.jar  
ru.mai.dep806.bigdata.mr.FilterPost /user/stud/stackoverflow/landing/Posts /user/  
stud/liya/mapReduce_FilterPost
```

MapReduce Time: 1050002 msec

Запрос 2:

На шаге Map сделать сплит по тегам.

На шаге Reduce просуммировать `result` по годам и тегам.

```
hadoop jar myMapReduce-1.0-SNAPSHOT.jar ru.mai.dep806.bigdata.mr.TagSplit /  
user/stud/liya/mapReduce_FilterPost /user/stud/liya/mapReduce_TagSplit
```

Реализовано в классе `TagSplit.class`

MapReduce Time: 52327 msec

Запрос 3:

Отсеиваем те теги, которые не вошли 2020.

Реализовано в классе `FilterByList`

```
hadoop jar myMapReduce-1.0-SNAPSHOT.jar  
ru.mai.dep806.bigdata.mr.FilterByList /user/stud/liya/mapReduce_TagSplit /user/  
stud/liya/mapReduce_FilterByList
```

MapReduce Time: 45206 msec

Запрос 4:

Аналогичный запрос 5-6 в Hive, где требуется определить коэффициенты А,В,С,Д,Е и построить функцию $f(year) = a * (year - 2008) + b$.

Реализовано в классе CoefByTag

```
hadoop jar myMapReduce-1.0-SNAPSHOT.jar ru.mai.dep806.bigdata.mr.CoefByTag  
/user/stud/liya/mapReduce_FilterByList /user/stud/liya/mapReduce_CoefByTag
```

MapReduce Time: 62362 msec

Запрос 5:

Аналогичный запрос 7 в Hive. Требуется определить топ 10 трендовых технологий.

Реализовано в классе TopTrand

```
hadoop jar myMapReduce-1.0-SNAPSHOT.jar ru.mai.dep806.bigdata.mr.TopTrand /  
user/stud/liya/mapReduce_CoefByTag /user/stud/liya/mapReduce_TopTrand
```

MapReduce Time: 43247 msec

**Результаты после запроса 5 сохранены в /user/stud/liya/
mapReduce_TopTrand/part-r-00000**

angular	2145.576082142853
typescript	996.9441111111113
reactjs	881.1713999999979
android-studio	868.4499939393937
android	785.6721076923131
ios	721.4288076923067
python-3.x	699.7711615384621
xcode	380.359842307692
jenkins	289.3589342857143
pip	263.3420712121217

Total MapReduce Time: 20 minutes 53 seconds 144 msec