

CprE419 Lab 6
Neha Maddali

Experiment 1:

Move github.csv to hadoop: *hadoop fs -put ~/Downloads/github.csv /data*

spark-submit --class Lab6_Exp1 ~/Downloads/Lab6_Exp1.jar /data/github.csv /lab6/exp1/output

Merge output files: *hadoop fs -cat /lab6/exp1/output/* | hadoop fs -put - /lab6/exp1/output.txt*

Check output folder: *hadoop fs -ls /lab6/exp1*

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /lab6/exp1/output.txt
2023-04-09 11:26:02,521 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Java 462182 elastic/elasticsearch 15698
Ruby 363801 rails/rails 29825
Python 331883 jakubroztocil/httpie 21283
PHP 273999 zurb/foundation 22636
C++ 159831 nwjs/nw.js 27350
C 145354 neovim/neovim 17395
C# 116155 dotnet/corefx 9176
```

Experiment 2:

Move patents_small.txt to hadoop: *hadoop fs -put ~/Downloads/patents_small.txt /data*

spark-submit --class Lab6_Exp2 ~/Downloads/Lab6_Exp2.jar /data/patents_small.txt

/lab6/exp2/output

Merge output files: *hadoop fs -cat /lab6/exp2/output/* | hadoop fs -put - /lab6/exp1/output.txt*

Check output folder: *hadoop fs -ls /lab6/exp2*

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /lab6/exp2/output.txt
2023-04-09 11:37:07,876 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
851768
```