

CprE419 Lab 5
Neha Maddali

Experiment 1:

spark-submit --class WordCountSpark ~/Downloads/WordCountSpark_Exp1.jar

/data/shakespeare /labSpark/output

Check output folder: `hadoop fs -ls /labSpark/output`

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /labSpark/output/part-00000 | head -n 10
2023-04-02 11:27:09,157 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localhostTrusted = false, remoteHostTrusted = false
(81827,,)
(36514,.)
(35942,)
(23272,the)
(20041,I)
(17274,;)
(16817,and)
(15506,to)
(15037,of)
(12361,you)
```

Experiment 2:

Move ip_trace and raw_block to hadoop: `hadoop fs -put ~/Downloads/ip_trace /data`

`hadoop fs -put ~/Downloads/raw_block /data`

spark-submit --class Lab5_Exp2 ~/Downloads/Lab5_Exp2.jar /data/ip_trace /data/raw_block

/lab5/output/parta /lab5/output/partb

PART A:

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /lab5/output/parta/part-00000 | head -n 10
2023-04-02 12:08:02,900 INFO sasl.SaslDataTransferClient: SASL encryption t
rust check: localhostTrusted = false, remoteHostTrusted = false
15:33:22.082247 2828 10.31.24.241.49231 129.186.140.200.53: Blocked
11:36:57.966302 400543 10.31.27.121.5353 224.0.0.251.5353: Blocked
19:54:27.197745 573151 10.33.5.84.5353 224.0.0.251.5353: Blocked
16:06:26.262910 43399 10.31.27.56.5353 224.0.0.251.5353: Blocked
20:04:48.002844 581541 10.31.5.228.5353 224.0.0.251.5353: Blocked
06:45:44.858596 366184 17.134.127.97.443 192.168.1.4.63890: Blocked
18:55:49.666388 257529 10.31.5.214.56417 224.0.0.252.5355: Blocked
19:42:15.717182 304364 151.101.148.133.443 10.31.4.211.63522: Blocked
16:51:02.568905 109349 13.249.140.62.443 10.31.24.241.63030: Blocked
15:36:59.636708 6611 10.33.15.125.5353 224.0.0.251.5353: Blocked
```

PART B

```
cpre419@cpre419-VirtualBox:~/hadoop/sbin$ hadoop fs -cat /lab5/output/partb
/part-00000 | head -n 10
2023-04-02 12:10:38,578 INFO sasl.SaslDataTransferClient: SASL encryption t
rust check: localhostTrusted = false, remoteHostTrusted = false
(73223,13.249.140.62.443)
(46495,23.35.205.78.443)
(18179,10.31.24.241.64121)
(11473,151.101.148.133.443)
(10590,130.211.46.84.443)
(10361,10.31.24.241.63030)
(3864,100.24.180.47.443)
(3809,10.31.27.253.1985)
(3764,10.31.27.252.1985)
(3357,10.31.24.241.64037)
```