

CPR E 419 - Lab 5

Spark

1.

Below are screenshots of the commands run and successful completion of program execution.

```
cpre419@cpre419-VirtualBox:~$ spark-submit --class WordCountSparkDec ~/Downloads/WordCountSpDec.jar /data/shakespeare /lab5/exp1/output
2021-04-01 03:26:36,057 WARN util.Utils: Your hostname, cpre419-VirtualBox resolves to a loopback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
2021-04-01 03:26:36,059 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
2021-04-01 03:26:38,671 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-04-01 03:26:39,312 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers
2021-04-01 03:26:39,496 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (8192 MB per container)
2021-04-01 03:26:39,497 INFO yarn.Client: Will allocate AM container, with 5632 MB memory including 512 MB overhead
2021-04-01 03:26:39,498 INFO yarn.Client: Setting up container launch context for our AM
2021-04-01 03:26:39,500 INFO yarn.Client: Setting up the launch environment for our AM container
2021-04-01 03:26:39,532 INFO yarn.Client: Preparing resources for our AM container
2021-04-01 03:26:39,595 WARN yarn.Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.
2021-04-01 03:26:41,923 INFO yarn.Client: Uploading resource file:/tmp/spark-b08dfc54-88fa-47be-800e-8234a478bea7/_spark_libs_2757420615080469347.zip -> hdfs://localhost:9000/user/cpre419/.sparkStaging/application_1616204106748_0005/_spark_libs_2757420615080469347.zip
2021-04-01 03:26:45,197 INFO yarn.Client: Uploading resource file:/home/cpre419/Downloads/WordCountSpDec.jar -> hdfs://localhost:9000/user/cpre419/.sparkStaging/application_1616204106748_0005/WordCountSpDec.jar
```

```
client token: N/A
diagnostics: N/A
ApplicationMaster host: 10.0.2.15
ApplicationMaster RPC port: 44907
queue: default
start time: 1617265607533
final status: SUCCEEDED
tracking URL: http://cpre419-VirtualBox:8088/proxy/application_1616204106748_0005/
user: cpre419
2021-04-01 03:27:34,809 INFO util.ShutdownHookManager: Shutdown hook called
2021-04-01 03:27:34,818 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-b0193d3a-d0f9-4445-8acd-c35e1bc91b79
2021-04-01 03:27:34,824 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-b08dfc54-88fa-47be-800e-8234a478bea7
```

The first 10 words obtained in decreasing order are as follows:

```
(81827,,)
(36514,.)
(35942,)
(23272,the)
(20041,I)
(17274,;)
(16817,and)
(15506,to)
(15037,of)
(12361,you)
```

2.

Below are screenshots of the commands run and successful completion of program execution.

```
cpred19@cpre419-VirtualBox:~$ spark-submit --class Firewall ~/Downloads/Firewallspark.jar /data/ip_trace/data/raw_block/lab5/exp2/firewall/output/lab5/exp2/blocked/output
2021-04-01 11:47:58,697 WARN util.Utils: Your hostname, cpre419-VirtualBox resolves to a loopback address: 127.0.0.1; using 10.0.2.15 instead (on interface enp0s3)
2021-04-01 11:47:58,700 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
2021-04-01 11:48:04,070 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-04-01 11:48:04,974 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers
2021-04-01 11:48:05,323 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (8192 MB per container)
2021-04-01 11:48:05,325 INFO yarn.Client: Will allocate AM container, with 5632 MB memory including 512 MB overhead
2021-04-01 11:48:05,326 INFO yarn.Client: Setting up container launch context for our AM
2021-04-01 11:48:05,328 INFO yarn.Client: Setting up the launch environment for our AM container
2021-04-01 11:48:05,390 INFO yarn.Client: Preparing resources for our AM container
2021-04-01 11:48:05,664 WARN yarn.Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.
2021-04-01 11:48:09,623 INFO yarn.Client: Uploading resource file:/tmp/spark-e34a50e1-9398-46b2-bde4-672f9885f932/_spark_libs_7714387015027930143.zip -> hdfs://localhost:9000/user/cpre419/.sparkStaging/application_1616204106748_0006/_spark_libs_7714387015027930143.zip
2021-04-01 11:48:14,798 INFO yarn.Client: Uploading resource file:/home/cpre419/Downloads/Firewallspark.jar -> hdfs://localhost:9000/user/cpre419/.sparkStaging/application_1616204106748_0006/Firewallspark.jar
2021-04-01 11:48:15,567 INFO yarn.Client: Uploading resource file:/tmp/spark-e34a50e1-9398-46b2-bde4-672f9885f932/_spark_conf_1204437129054564742.zip -> hdfs://localhost:9000/user/cpre419/.sparkStaging/application_1616204106748_0006/_spark_conf_1204437129054564742.zip
2021-04-01 11:48:15,673 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.closeResponder(DFSOutputStream.java:609)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.endBlock(DFSOutputStream.java:370)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.run(DFSOutputStream.java:546)
```

```
client token: N/A
diagnostics: N/A
ApplicationMaster host: 10.0.2.15
ApplicationMaster RPC port: 42799
queue: default
start time: 1617295701885
final status: SUCCEEDED
tracking URL: http://cpre419-VirtualBox:8088/proxy/application_1616204106748_0006/
user: cpre419
2021-04-01 11:51:08,018 INFO util.ShutdownHookManager: Shutdown hook called
2021-04-01 11:51:08,023 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-969f8e13-c310-40d2-b312-69cbc02017e4
2021-04-01 11:51:08,030 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-e34a50e1-9398-46b2-bde4-672f9885f932
```

Firewall log file output is of the following format:

```
15:28:46.539147 0 10.31.27.252 224.0.0.102 Blocked
15:28:47.479137 1 10.31.26.186 224.0.0.251 Blocked
15:28:47.479217 2 10.31.26.216 224.0.0.251 Blocked
15:28:48.105882 3 10.31.27.253 224.0.0.102 Blocked
15:28:48.628287 5 10.10.251.232 224.0.0.1: Blocked
15:28:49.672578 8 10.31.26.216 224.0.0.251 Blocked
15:28:50.194885 10 10.31.24.175 224.0.0.251 Blocked
15:28:50.299328 11 10.33.14.68 224.0.0.251 Blocked
15:28:50.403757 12 10.33.14.68 224.0.0.251 Blocked
15:28:50.508240 13 10.31.26.216 224.0.0.251 Blocked
15:28:50.717141 14 10.31.27.253 224.0.0.102 Blocked
```

The blocked IPs in decreasing order of sorting are:

```
(73223,13.249.140.62)
(65013,10.31.24.241)
(46495,23.35.205.78)
(12103,10.31.4.211)
(11473,151.101.148.133)
(11470,192.168.1.4)
(10590,130.211.46.84)
(8690,10.31.4.188)
(4959,10.31.6.140)
(4225,10.33.15.125)
```