

How To Execute Hadoop Code:

1. Start dfs and yarn

```
hadoop@tom-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [tom-VirtualBox]
2021-12-09 21:02:00,175 WARN util.NativeCodeLoader: Unable
sses where applicable
hadoop@tom-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@tom-VirtualBox:~/hadoop-3.2.1/sbin$ jps
2145 DataNode
2321 SecondaryNameNode
2567 ResourceManager
3049 Jps
2010 NameNode
2699 NodeManager
```

2. Set Hadoop classpath

```
hadoop@tom-VirtualBox:/home/tom/Desktop/WordCountTutorial$ export HADOOP_CLASSPATH=/usr/lib/jvm/java-8-openjdk-amd64/lib/tools.jar
```

3. Delete previous output (if necessary)

```
hadoop@tom-VirtualBox:/home/tom/Desktop/WordCountTutorial$ hdfs dfs -rm -r TitanicKMeans/output
2021-12-09 21:05:23,160 WARN util.NativeCodeLoader: Unable to load native-hadoop library for yo
sses where applicable
Deleted TitanicKMeans/output
```

4. Run the job

```
hadoop@tom-VirtualBox:/home/tom/Desktop/WordCountTutorial$ hadoop jar SSA1.jar TitanicKMeans TitanicKMeans/input TitanicKMeans/output
```

```
2021-12-09 21:07:22,940 INFO mapreduce.Job: Running job: job_1639101745511_0001
2021-12-09 21:07:34,296 INFO mapreduce.Job: Job job_1639101745511_0001 running in uber mode : false
2021-12-09 21:07:34,297 INFO mapreduce.Job: map 0% reduce 0%
2021-12-09 21:07:41,418 INFO mapreduce.Job: map 100% reduce 0%
2021-12-09 21:07:48,473 INFO mapreduce.Job: map 100% reduce 100%
2021-12-09 21:07:48,510 INFO mapreduce.Job: Job job_1639101745511_0001 completed successfully
2021-12-09 21:07:48,694 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=390
    FILE: Number of bytes written=451267
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=227
    HDFS: Number of bytes written=360
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
```

```

HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4141
  Total time spent by all reduces in occupied slots (ms)=4406
  Total time spent by all map tasks (ms)=4141
  Total time spent by all reduce tasks (ms)=4406
  Total vcore-milliseconds taken by all map tasks=4141
  Total vcore-milliseconds taken by all reduce tasks=4406
  Total megabyte-milliseconds taken by all map tasks=4240384
  Total megabyte-milliseconds taken by all reduce tasks=4511744

```

5. Display output

```

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=102
File Output Format Counters
  Bytes Written=360
hadoop@tom-VirtualBox: /home/tom/Desktop/WordCountTutorial$ hdfs dfs -cat TitanicKMeans/output/part-r-00000
2021-12-09 21:08:24,145 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cla
sses where applicable
2021-12-09 21:08:25,492 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted =
false
0 [[0.42424242, 0.58441558, 14.64575758], [0.37162162, 0.67567568, 51.20945946], [0.36914062, 0.66796875, 30.27293716]]

```