SPARK RUN STEPS

1) export AWS_ACCESS_KEY_ID=
2) export AWS_SECRET_ACCESS_KEY=
3) cd spark/ec2
4) Below command to launch the spark cluster with 16 slaves and one master

   ./spark-ec2 --key-pair=hadoopec2cluster --identity-file=/home/ubuntu/hadoopec2cluster.pem --region=us-east-1 --slaves=16 --instance-type=c3.large --ebs-vol-size=750 --ami=ami-877142ed --spot-price=0.04 launch spark

5) Login into spark cluster
   ./spark-ec2 -k hadoopec2cluster -i /home/ubuntu/hadoopec2cluster.pem -r us-east-1 login Spark
6) Deploy the spark jar file to run sorting
7) Login to the master node using winscp
   ~/spark-ec2/copy-dir /root/SparkSort-1.0-SNAPSHOT.jar
8) cd ephemeral-hdfs/bin
9) ./hadoop dfs -copyFromLocal /datafile.txt /datain
10) Run the below command to submit the job
    ./spark-submit --class SparkSort --master local[8] --executor-cores 2 --conf spark.driver.memory=2g --conf spark.executor.memory=2g /root/SparkSort-1.0-SNAPSHOT.jar /sorted100gb /sortedout
11) cd ephemeral-hdfs/bin
12) ./hadoop dfs -copyToLocal /dataout /sortedout