CS553 Cloud Computing

Programming Assignment #2

Manual

Pranitha Nagavelli(A20345406)

Compiling in Linux(ubuntu)OS & AWS Instance:

GenSort:
Step1
- Download gensort-linux-1.5.tar.gz and untar it in the instance

Step2:
- Go to either 64 or 32 folder depending on the system

Step3:
- ./gensort -a <sizeoffile>1000000 <filename along with locaton> /roor/cache/input. Using this command produce dataset needed.

Shared-Memory Sort:
Step1
- Send the .java file into the instance using SCP command from the local machine

Step2:
- Run $javac sharedSort.java

Step3:
- Run command $java sharedSort inputfile outputfile

Hadoop:
Step1
- Send the Hadoop_Sort and Pem file into the instance using SCP command from the local machine

Step2:
- Get the data into hdfs using following command: bin/hdfs dfs -put /home/ec2-user /cache/data/dataset /dataset

Step3:
- Run the Hadoop jar file Command:bin/hdfs jar /home/ec2-user/Hadoop_Sort.jar /dataset /output

Step4:
- Get the data from hdfs using following command: bin/hdfs dfs -get /output   /home/ec2-user/cache/data/dataset

Spark:
Hadoop:
Step1
- Send the SparkSort.jar and Pem file into the instance using SCP command from the local machine

Step2:
- Get the data into hdfs using following command: bin/hdfs dfs -put /home/ec2-user /cache/data/dataset /dataset

Step3:

- Run the Hadoop jar file Command: bin/spark-submit --class com.aamend.spark.SparkSort.Sort --master spark://52.90.143.190 /root/SparkSort.jar /dataset /output

Step4:

- Get the data from hdfs using following command: bin/hdfs dfs -get /output   /home/ec2-user/cache/data/dataset