The bold-italized texts are commands.

***start-all.sh***

***hdfs dfs -copyFromLocal ./<source path> /<destination path>***

<destination path> is the project folder created

To verify the copy: it list all files in the directory

***hdfs dfs -ls /<destination path>/***

Write a mapper.py and reducer.py. To run these python files we need streaming APIs of Hadoop

API Path : hadoop-3.1.2/share/hadoop/tools/lib/hadoop-streaming-3.1.2.jar

Mapper path: mapper.py in <destination path>

Reducer path: reducer.py in <destination path>

Results path: folder to store outputs in <destination path>

***hadoop jar <API Path> -file <Mapper path> -mapper <Mapper path> -file <Reducer path> -reducer <Reducer path> -input <destination path> -output <Results path>***

To check the results: we need to know the file name of the file created

***hdfs dfs -ls /<Results path>/***

The file created might be in the form of “part-00000” and to check result

***hdfs dfs -cat /<Results path>/***

Now to copy the output file from hdfs to local directory

***hdfs dfs -copyToLocal /<Results path>/filename ./<desired path>/<filename>.txt***

Use cat to verify the text file in the local system.

***cat /<desired path>/<filename>.txt***

Most important is to stop the services of hadoop

***stop-all.sh***