

Manual

1. First you must go to hadoop direction by using this command in terminal
`cd $HADOOP_HOME`
2. Start hadoop by using this command
`sbin/start-all.sh`
3. make direction in hdfs (hadoop distributed file system) by using this command
`hadoop dfs -mkdir -p ____ => the path you want to make`
4. put your files you will work on it in the direction you make by using this command
`hadoop dfs -put the file path the path you make in last step`
(Note)between the file path and the path in hdfs you must put a space
(Note)you have to files you must this command two times to put them
5. Now we can run the job by using this command
`hadoop jar /usr/local/hadoop-2.8.1/share/hadoop/tools/lib/hadoop-streaming-2.8.1.jar \`
-input the path of file you work in the hadoop dfs \
-output a direction you make and should not exist \
-file the mapper file direction \
-mapper the mapper direction \
-file the reducer direction \
-reducer the reducer direction
(Note) you have two file input enter the files paths separated by space
6. Now we have output file after run the last step it will be the the input for this step
`hadoop jar /usr/local/hadoop-2.8.1/share/hadoop/tools/lib/hadoop-streaming-2.8.1.jar \`
-input the path of the output in the last step \
-output a direction you make and should not exist \
-file the mapper2 file direction \
-mapper the mapper2 direction \
-file the reducer2 direction \
-reducer the reducer2 direction
7. open you browser and enter localhost:50070
8. go to Utilities => browse the file system and search for the output file from the last command and download (part-00000)

Test code

```
cd $HADOOP_HOME
sbin/start-all.sh
hadoop dfs -mkdir -p /user/abdallah/task2
hadoop dfs -put /home/abdallah/PycharmProjects/osTask2/user.txt /user/abdallah/task2
hadoop dfs -put /home/abdallah/PycharmProjects/osTask2/item.txt /user/abdallah/task2
hadoop jar /usr/local/hadoop-2.8.1/share/hadoop/tools/lib/hadoop-streaming-2.8.1.jar \
-input /user/abdallah/task2/user.txt /user/abdallah/task2/item.txt \
-output /user/abdallah/output \
-file /home/abdallah/PycharmProjects/osTask2/mapper.py \
-mapper /home/abdallah/PycharmProjects/osTask2/mapper.py \
-file /home/abdallah/PycharmProjects/osTask2/reducer.py \
-reducer /home/abdallah/PycharmProjects/osTask2/reducer.py
hadoop jar /usr/local/hadoop-2.8.1/share/hadoop/tools/lib/hadoop-streaming-2.8.1.jar
-input /user/abdallah/output/part-00000
-output /user/abdallah/output2
-file /home/abdallah/PycharmProjects/osTask2/mapper2.py
-mapper /home/abdallah/PycharmProjects/osTask2/mapper2.py
-file /home/abdallah/PycharmProjects/osTask2/reducer2.py
-reducer /home/abdallah/PycharmProjects/osTask2/reducer2.py
```