

Test

run code by python on terminal

```
cat user.txt item.txt | python mapper.py | sort -k1,1 | python reducer.py | python  
mapper2.py | sort -k1,1|python reducer2.py
```

(Note) you must use sort -k1,1 to sort your output to make your code run correctly

run code by hadoop

```
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
```