

Manual for Task #2

Prerequisites:

- This tutorial is developed on Ubuntu **operating System**.
- You should have **Hadoop (version 2.2.0)** used for this tutorial) already installed.
- You should have Java(**version 1.8.0** used for this tutorial) already installed on the system.

Steps to run the code:

Step 1) change directory to reach the path where hadoop installed

```
$ cd /usr/local/hadoop
```

Step 2)Start Hadoop

```
$HADOOP_HOME/sbin/start-dfs.sh
```

Step 3)make a directory on the hdfs

```
Hadoop hdfs dfs -mkdir transactions
```

Step 4)put the transaction & user files on this directory

```
Hadoop dfs -put ~/task2/job1/User.txt
```

```
Hadoop dfs -put  
~/task2/job1/Transaction.txt
```

Step 5)run the mapper & reducer on those file using hadoop streaming

```
hadoop jar /usr/local/hadoop-  
2.8.1/share/hadoop/tools/lib/hadoop-  
streaming-2.8.1.jar -input  
/usr/ahmedsaied/section3/* -output  
/usr/ahmedsaied/section3/output1 -mapper  
~/task2hadoop/job1/mapper.py -reducer  
~/task2hadoop/job1/reducer.py
```

Step 6) run the second mapreduce to get the expected output

```
hadoop jar /usr/local/hadoop-  
2.8.1/share/hadoop/tools/lib/hadoop-  
streaming-2.8.1.jar -input  
/usr/ahmedsaied/section3/output/part-  
00000 -output  
/usr/ahmedsaied/section3/output2 -mapper
```

```
~/task2hadoop/job1/mapper.py -reducer  
~/task2hadoop/job1/reducer.py
```