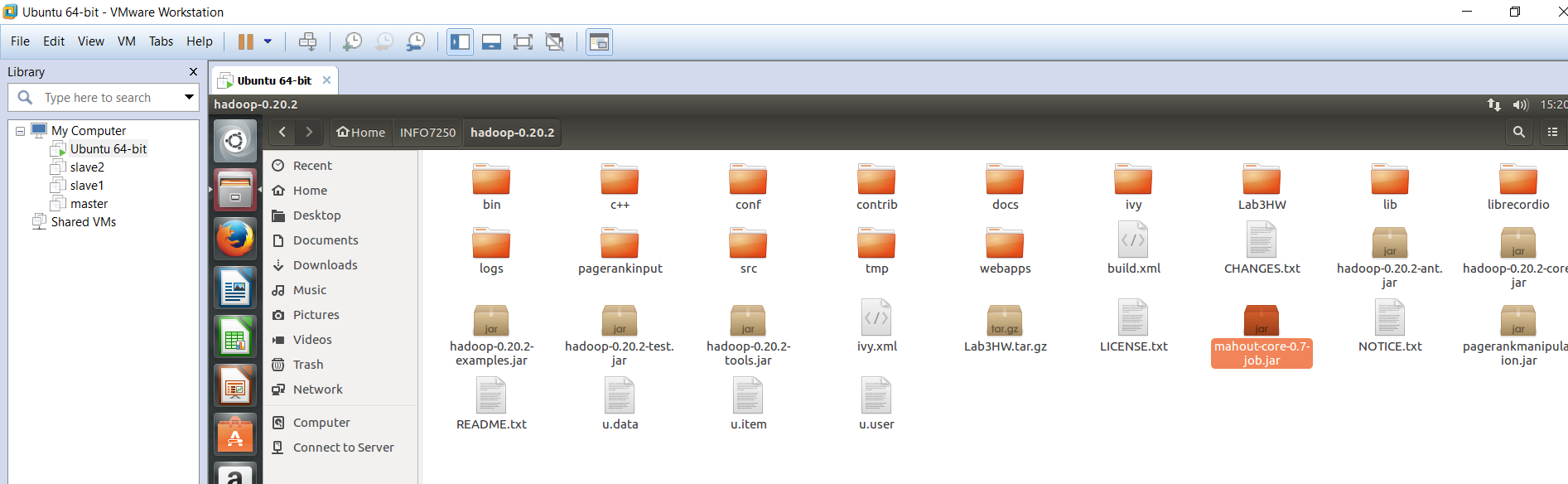
Question: To run Mahout on grouplens Movie dataset.

**Solution:**

1. Copied the mahout-core-0.7-job.jar to my Hadoop file location.

**Screenshot:**



I ran the job on pseudo distributed mode first by issuing the following command:

**Pseudo Mode:**

I moved the input file to HDFS by using the below commands:

**Commands:**

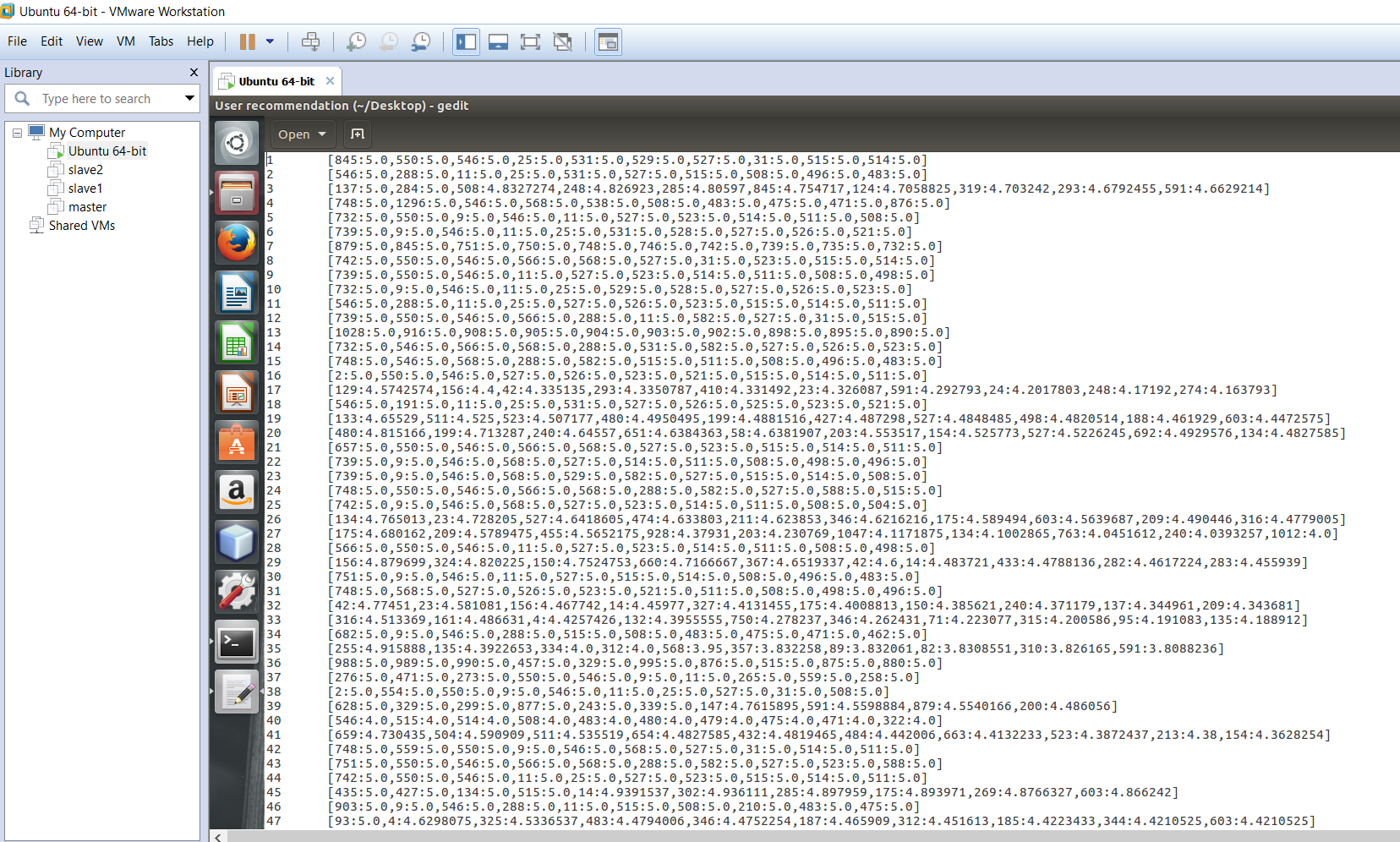
bin/hadoop fs -mkdir /Assignment7

bin/hadoop fs -mkdir /Assignment7/Input

bin/hadoop fs -copyFromLocal /home/shanmugasudan/INFO7250/hadoop-0.20.2/u.data /Assignment7/Input

**Command:** bin/hadoop jar /home/shanmugasudan/INFO7250/hadoop-0.20.2/mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s SIMILARITY\_COOCCURRENCE --input /Assignment7/Input/ --output /Assignment7/Output

**Output Screenshot:**

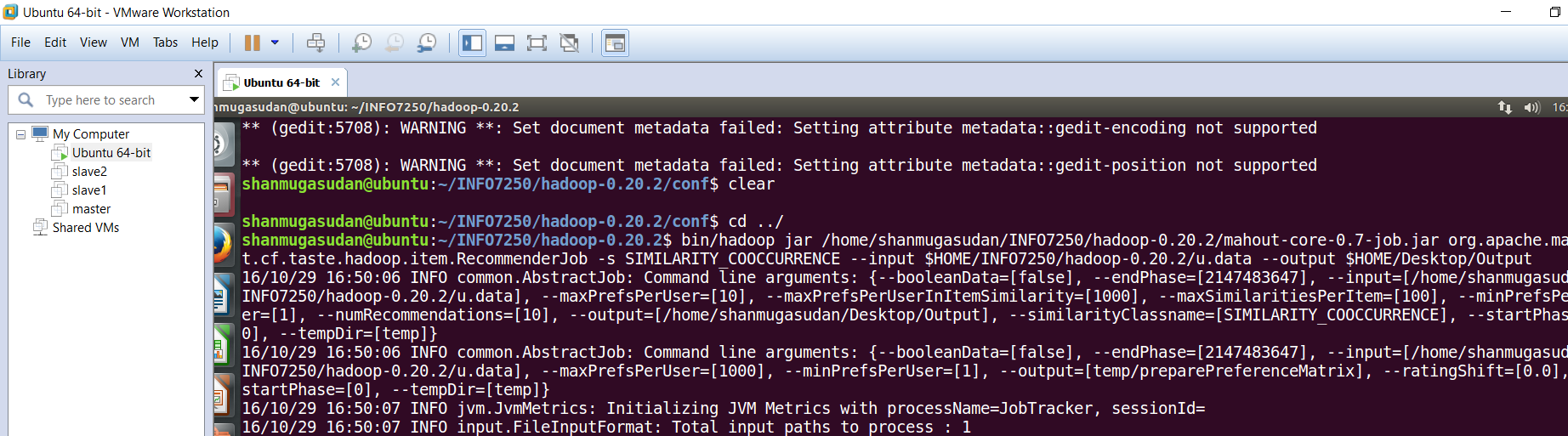


**Standalone Mode:**

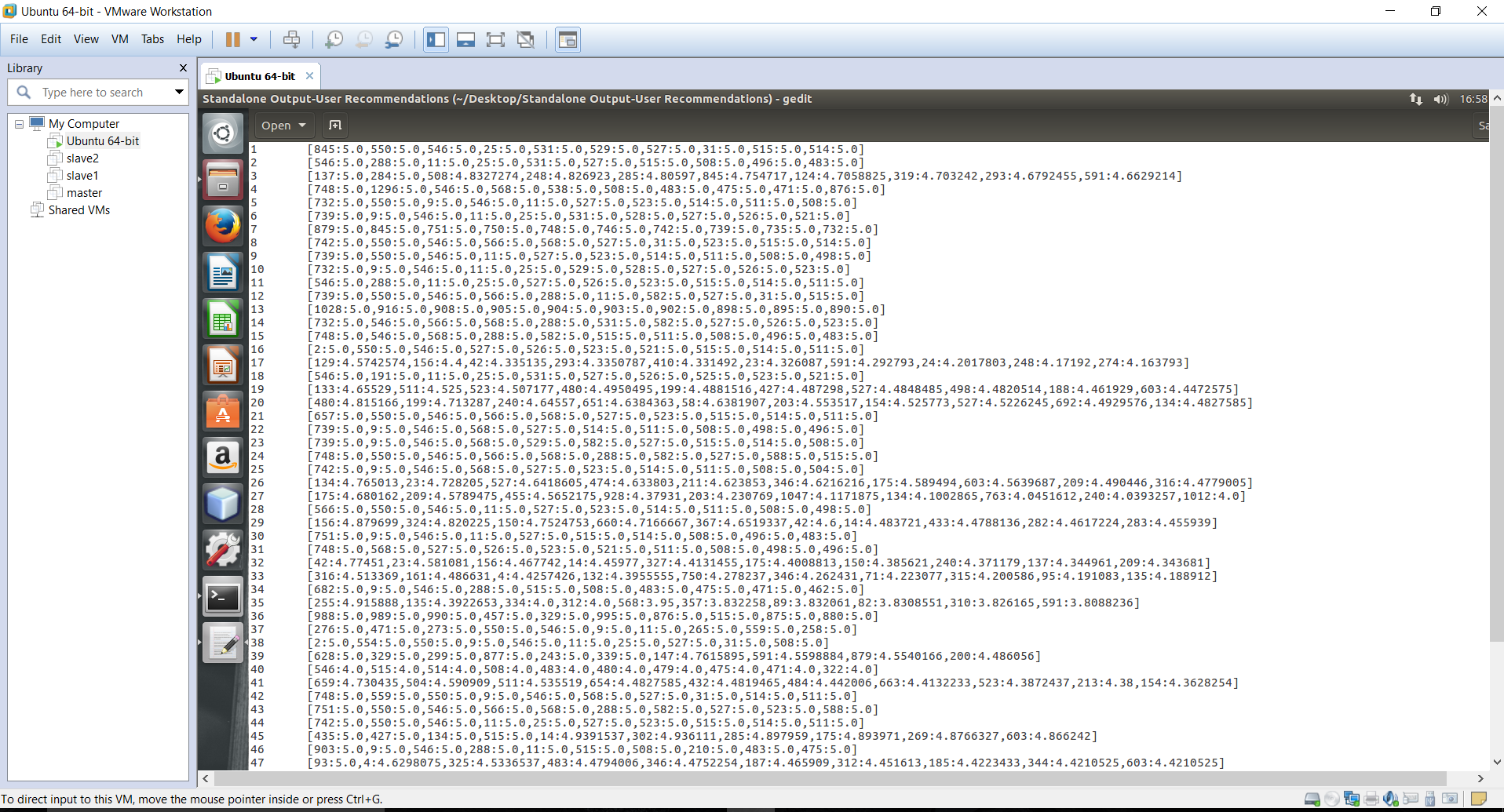
To run the code on standalone mode, I just commented the xml files (core-site, mapred-site, hdfs-site) and issued the below command:

**Command**: bin/hadoop jar /home/shanmugasudan/INFO7250/hadoop-0.20.2/mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s SIMILARITY\_COOCCURRENCE --input $HOME/INFO7250/hadoop-0.20.2/u.data --output $HOME/Desktop/Output

Screenshot sample:



Standalone output Screenshot:

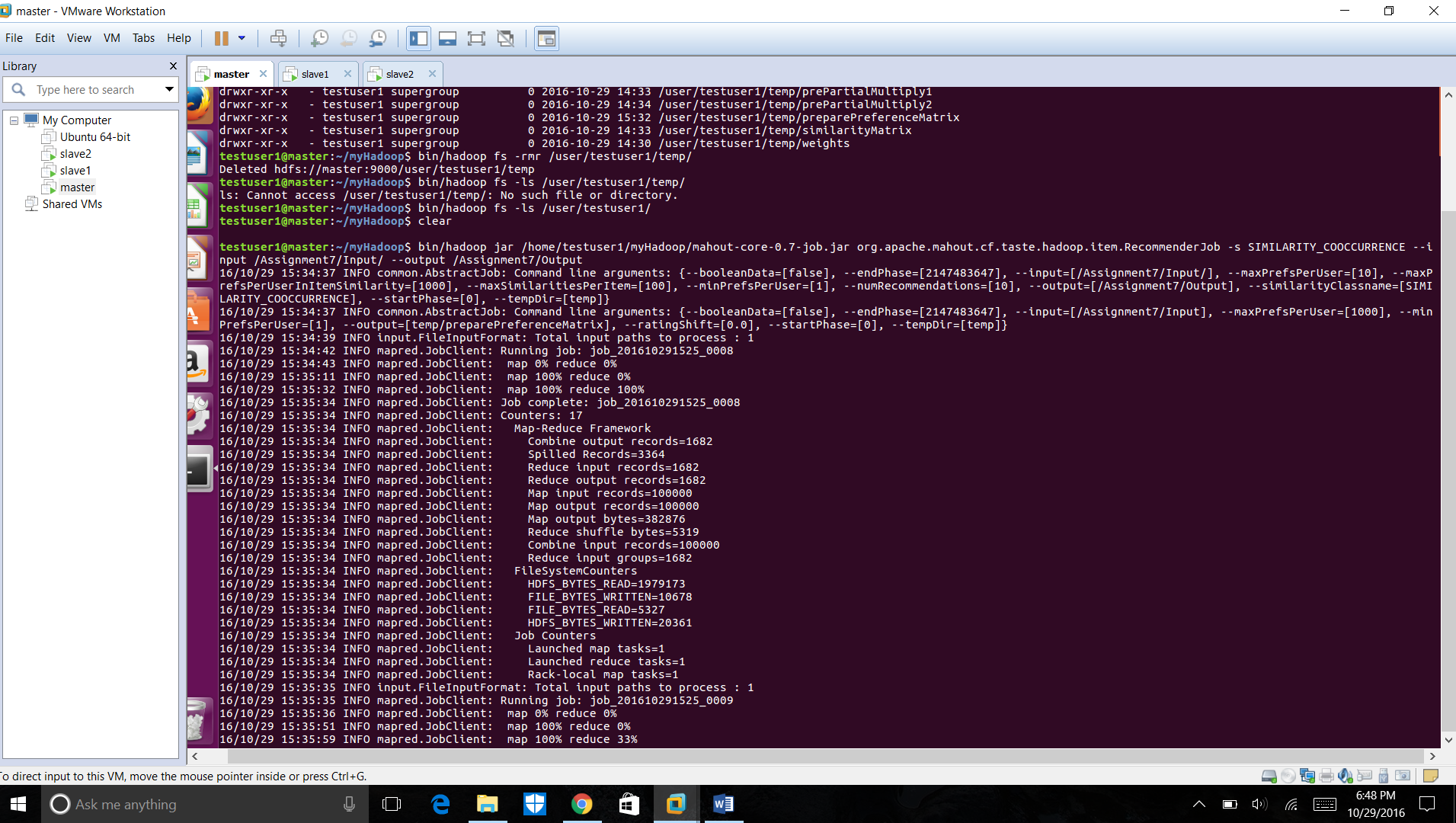


**Multi Node mode:**

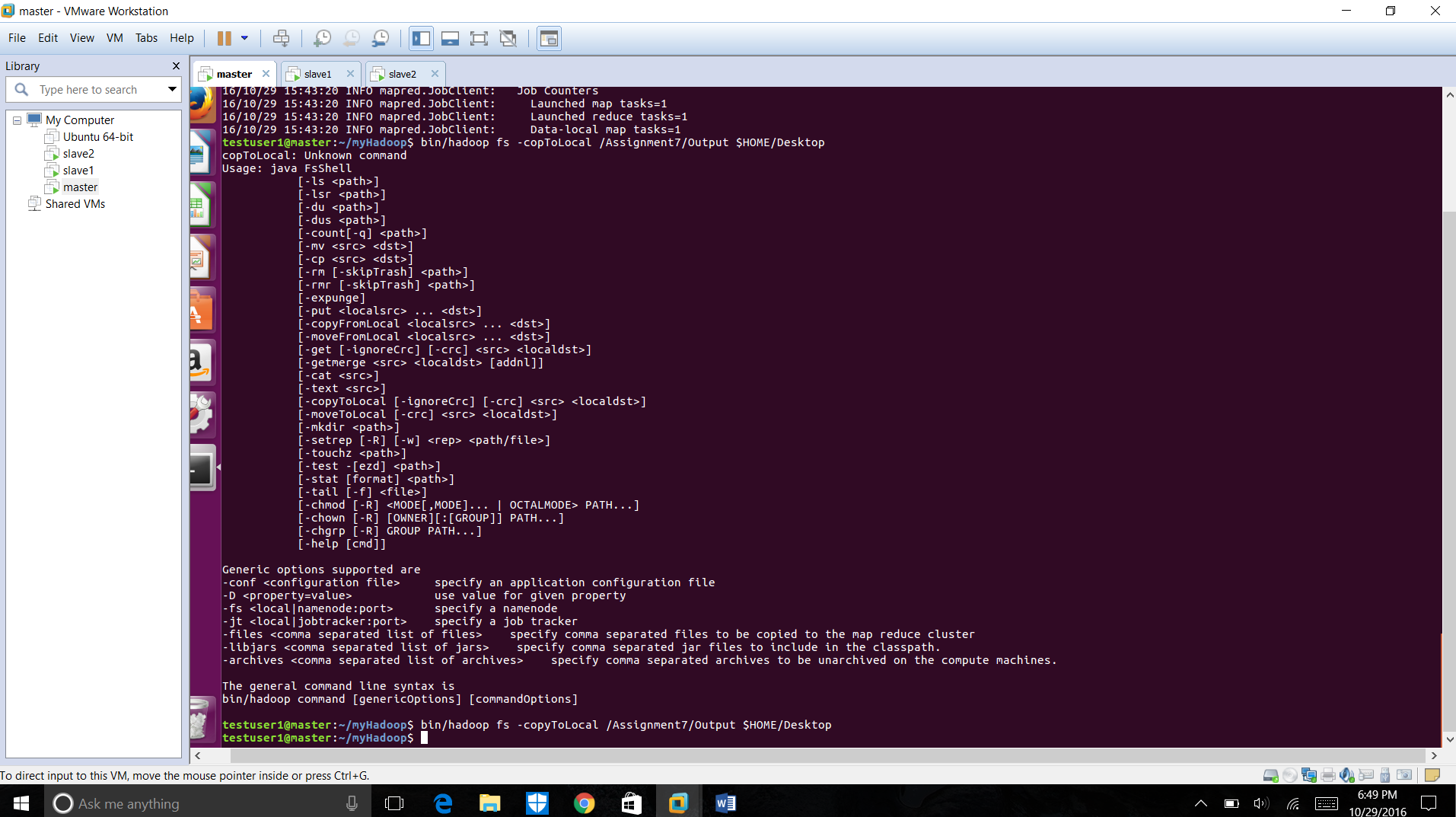
To run in multi node, I moved the mahout jar, input file to master node.

I kick started the Hadoop on master node.

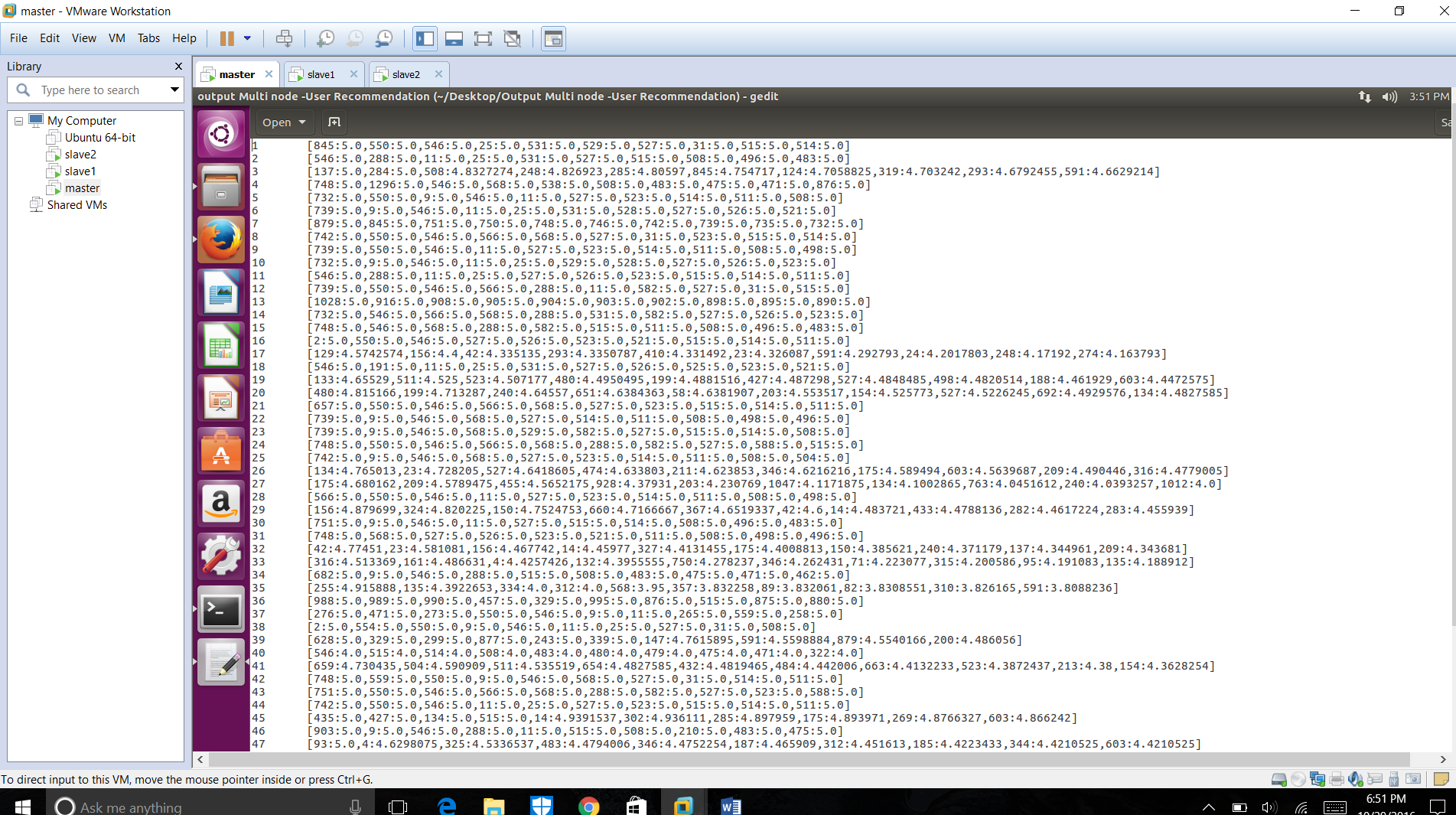
**Screenshot of my job submission on multi node:**



I ran the job and copied the file to my desktop,



**Multi node Output file:**



Note: During the multi node run, I was facing networking issues with the router, later I moved to location with better network connectivity.

**Single User Recommendation:**

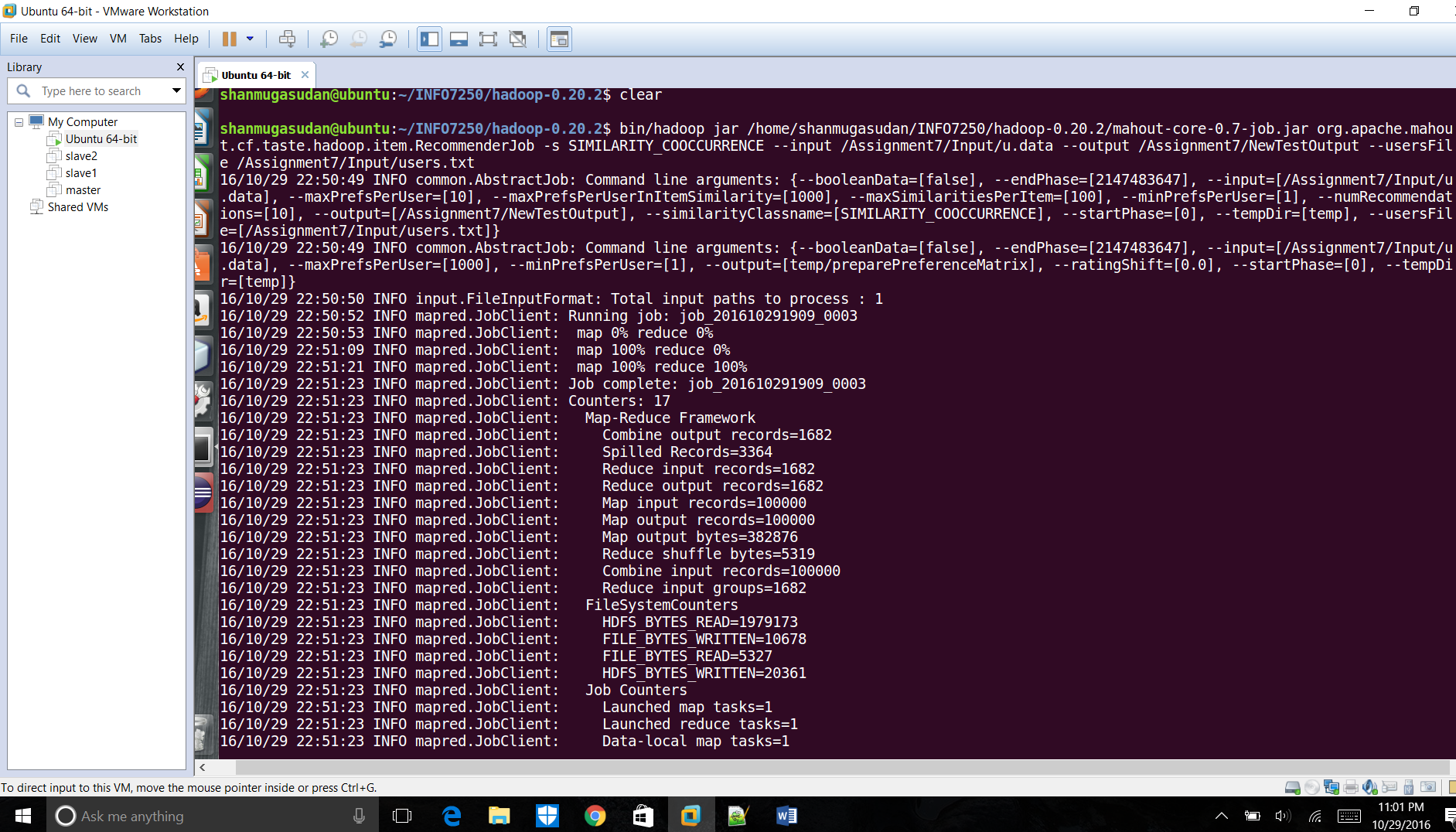
We just need to modify the command line input as

Command:

bin/hadoop jar /home/shanmugasudan/INFO7250/hadoop-0.20.2/mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s SIMILARITY\_COOCCURRENCE --input /Assignment7/Input/u.data --output /Assignment7/NewTestOutput ***--usersFile /Assignment7/Input/users.txt***

I have italicized the difference in command line for easy identification.

Screenshot:



Since the preference matrix is computed in the temp folder, I need to delete the files used in this folder before I could run the program for a single user.

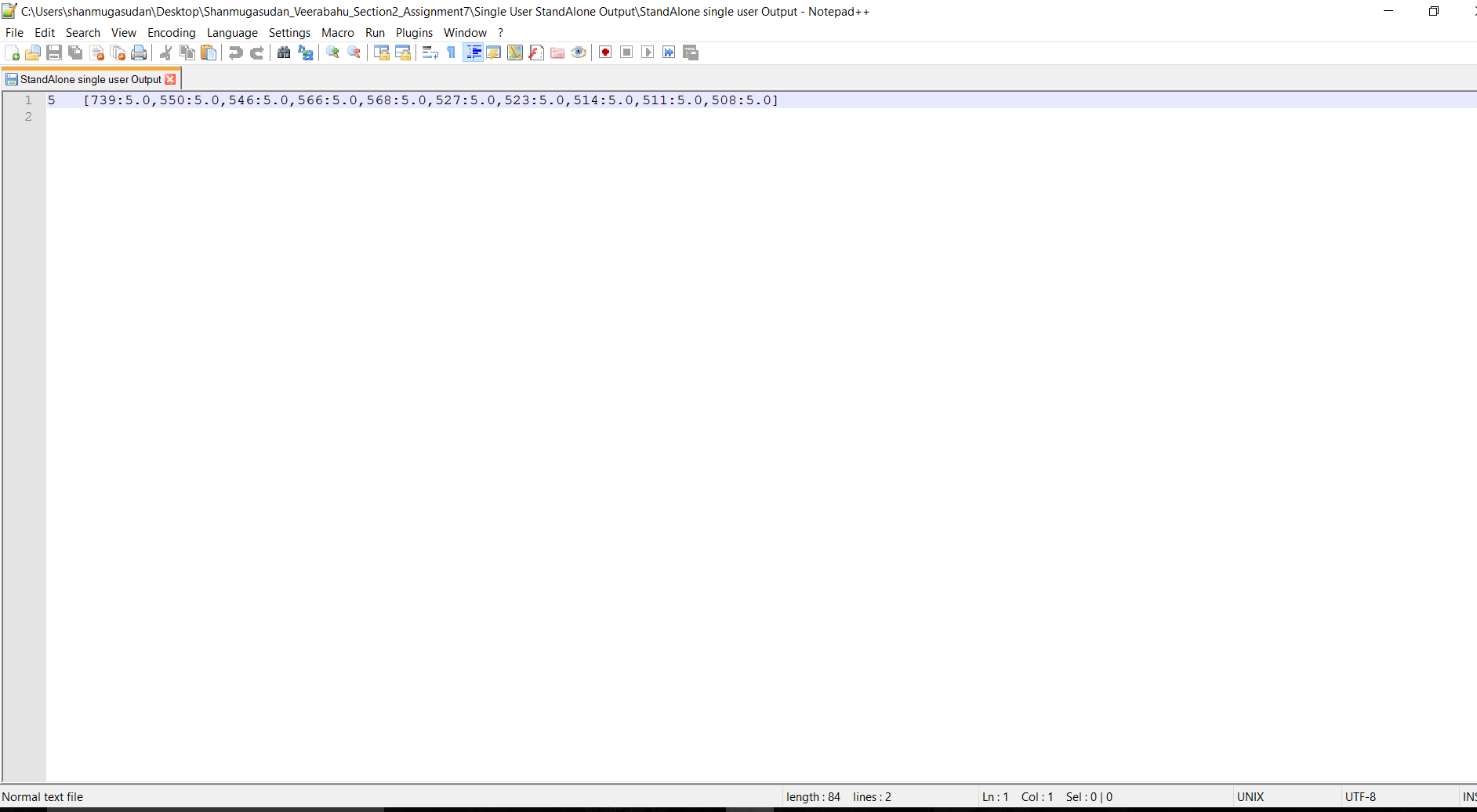
**Command:**

bin/hadoop fs -ls /home/shanmugasudan/INFO7250/hadoop-0.20.2/temp/

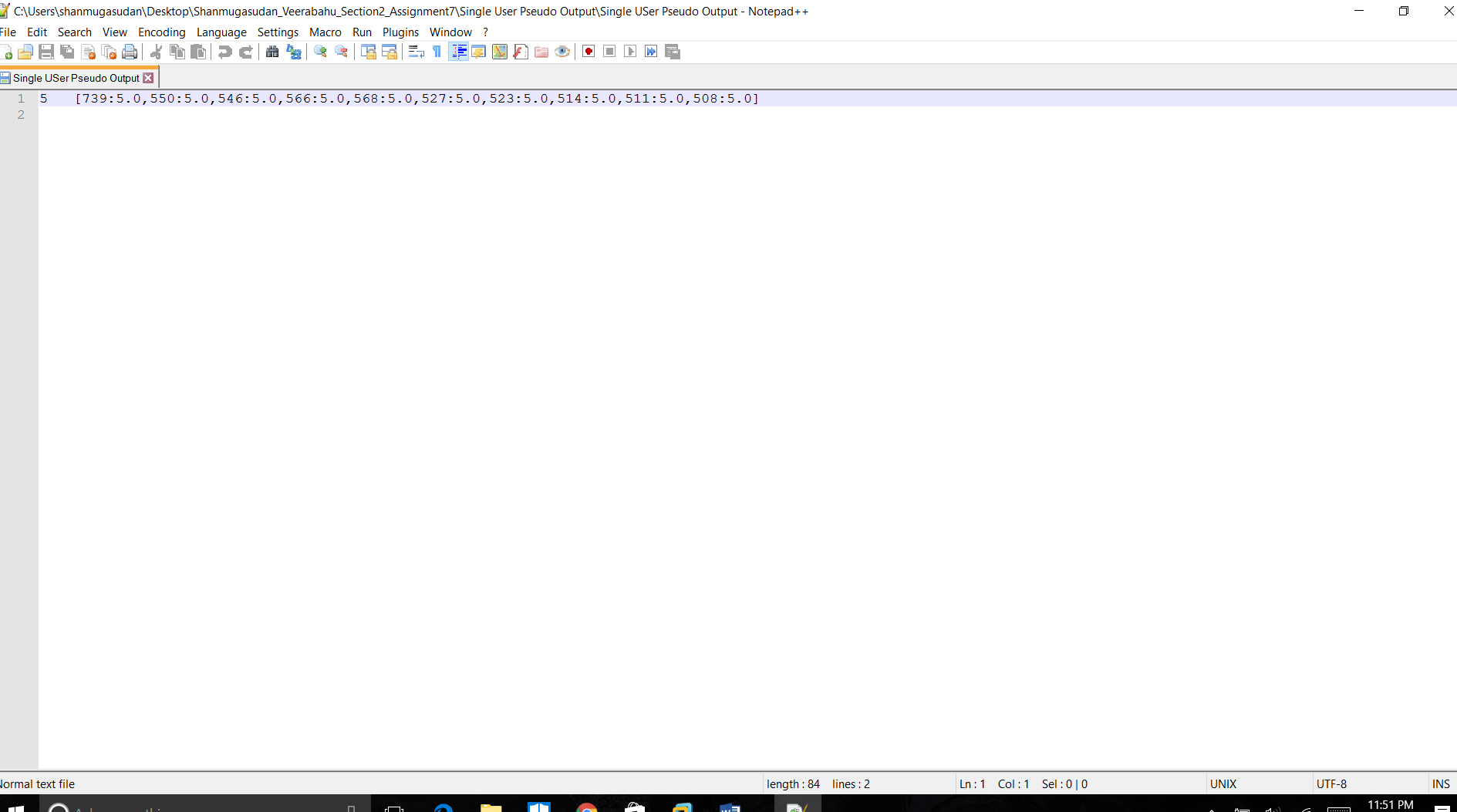
**Command used for Pseudo and Multi node mode:**

bin/hadoop jar /home/testuser1/myHadoop/mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s SIMILARITY\_COOCCURRENCE --input /Assignment7/Input/u.data --output /Assignment7/Output --usersFile /Assignment7/Input/users.txt

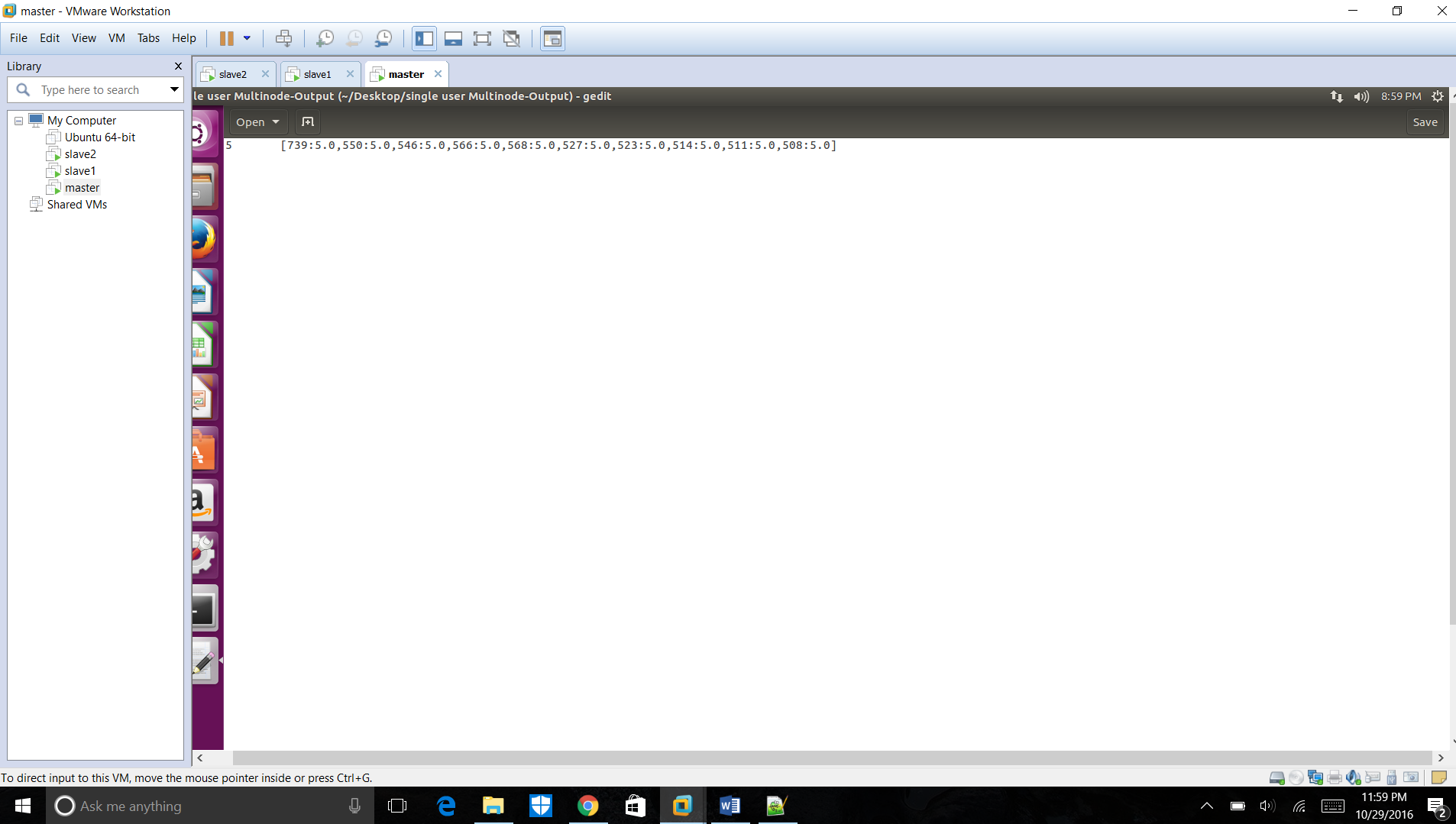
**Screenshots: Standalone mode (Taken after moving the file to my windows)**



**Pseudo mode: (Taken after moving the file to my windows)**



**Multi node mode:**



Actually, the recommender algorithm captures the ratings of all the movies for the users and so takes relatively the same time to calculate the output.

My First job to calculate recommendations for all users took me 15 mins approximately on a multimode cluster.

My Second job took me around 13 mins approximately to calculate the recommendations for single user.

When compared to single mode, pseudo mode, multi node cluster is very slow. It took a number of MR jobs to run and process the data compared to other modes.

When run on Amazon elastic MR, the cost to run the jobs would be high but the processing time may drop. ( Since it has got to deal with the Network bandwidth)

**Logistic classifier:**

My Dataset is a chess game dataset and has 38 features in total. I took the data from UCI Machine learning repository.

I ran the code provided for logistic classifier and achieved better accuracy. My Accuracy from winnow was 67% and for Logistic classifier, I have achieved an accuracy of 96% after L1 regularization and 100% after L2 regularization.

Screenshot of the result:

