**GroupLens 100k -Mahout**

Commands to use the mahout jar file and execute the jar on the Hadoop pseudo mode. We are sitting in a directory where Hadoop is.

1. We are going to use the mahout jar “**mahout-core-0.7-job.jar**” for running the recommendation for user based on the other users taste
2. We are going to use the **“scp” command** to send the file to the VM

***scp -r "inputGroupLens"*** [saksh@192.168.56.101:/home/saksh](mailto:saksh@192.168.56.101:/home/saksh)

1. We are going to use the Hadoop command to put the file from local toh HDFS, we are going to put the **u.data file** which has user id, movie id, ratings and timestamp, all are tab separated

***bin/hadoop fs -put "../Hadoop\_Workspace/inputGroupLens/u.data" recommenderInput/u.data***

1. We are going to put **users.txt file**  to the HDFS, this files consists of the users who recommendation is required to us. Our user file has 10 ids.

***bin/hadoop fs -put "../Hadoop\_Workspace/inputGroupLens/users.txt" recommenderInput/users.txt***

1. We are going to run the below command to get the recommendation for all the users present in the u.data file. The movies which are recommended to them are the one which they haven’t seen based on the similar users.

***hadoop jar mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -s SIMILARITY\_COOCCURRENCE --input recommenderInput/u.data --output recommenderInput/output***

the output will be stored in the output directory.

1. To get the recommendation for the specific users, we are going to run the below command

***bin/hadoop jar mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -Dmapred.input.dir=recommenderInput/u.data -Dmapred.output.dir=recommenderInput/OutputUser --usersFile recommenderInput/users.txt --booleanData false --similarityClassname SIMILARITY\_COOCCURRENCE***

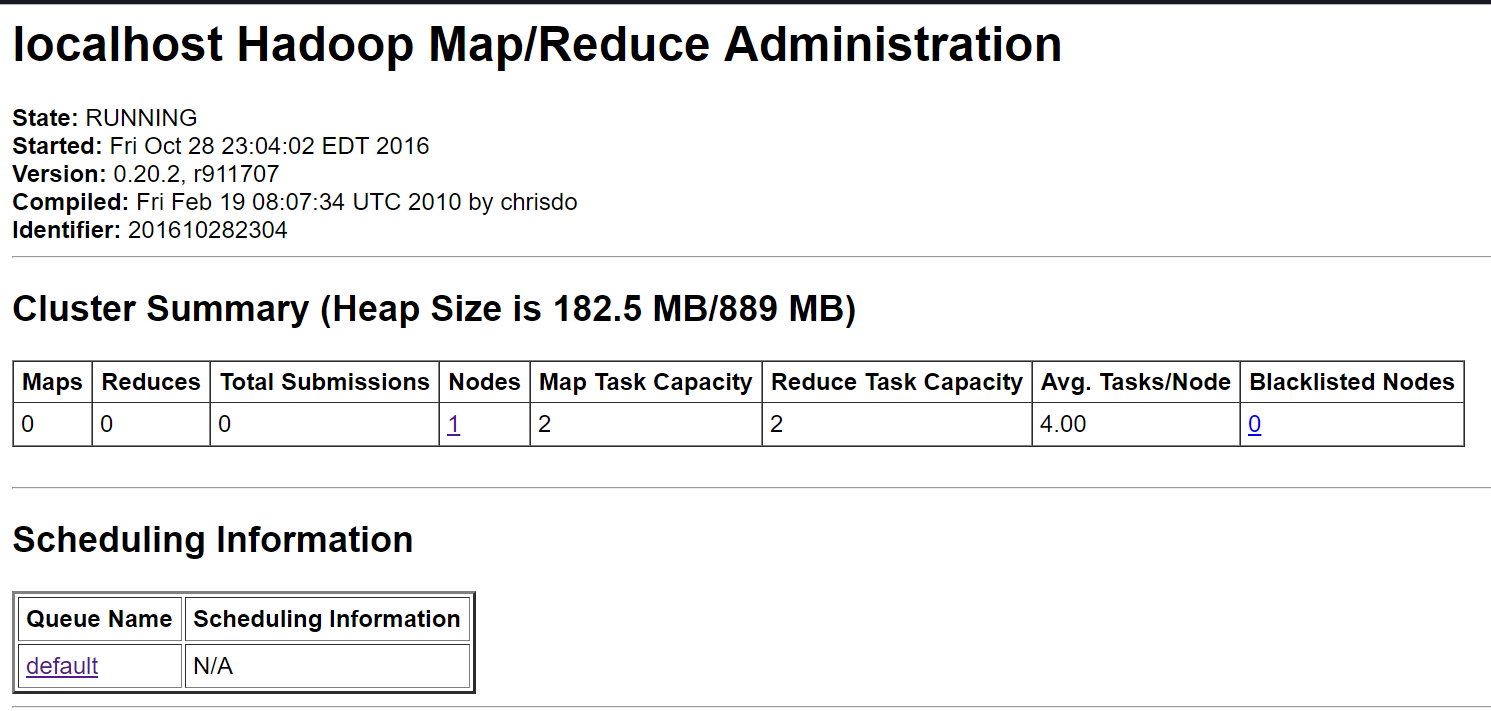
1. Command to get the recommendation for specific user in the **standalone mode** is

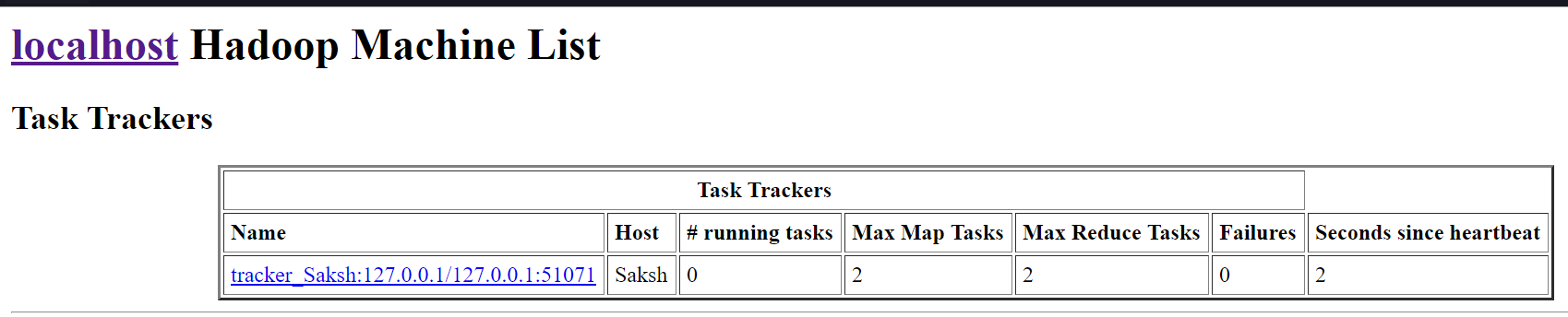
***bin/hadoop jar mahout-core-0.7-job.jar org.apache.mahout.cf.taste.hadoop.item.RecommenderJob -Dmapred.input.dir=../Hadoop\_Workspace/inputGroupLens/u.data -Dmapred.output.dir=../Hadoop\_Workspace/inputGroupLens/OutputUser --usersFile ../Hadoop\_Workspace/inputGroupLens/users.txt --booleanData false --similarityClassname SIMILARITY\_COOCCURRENCE***

1. **Pseudo mode execution Snapshots**

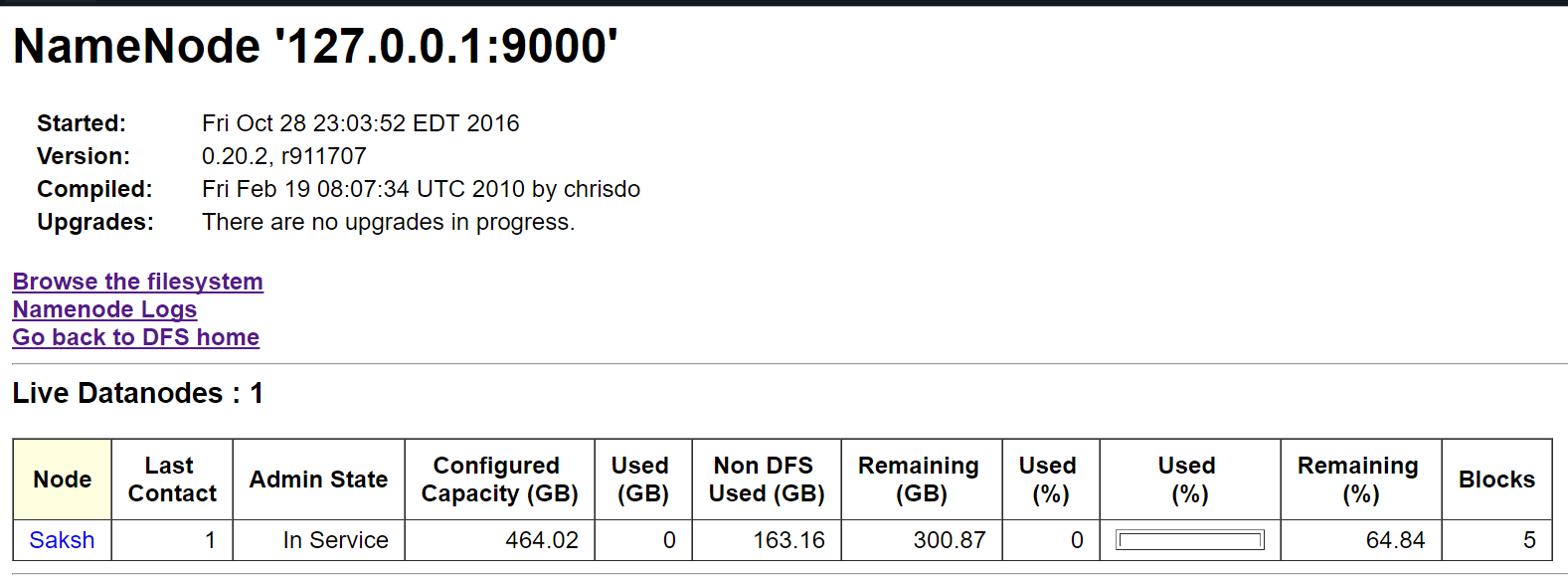
**Initial Setup**

1. **Localhost:50030**

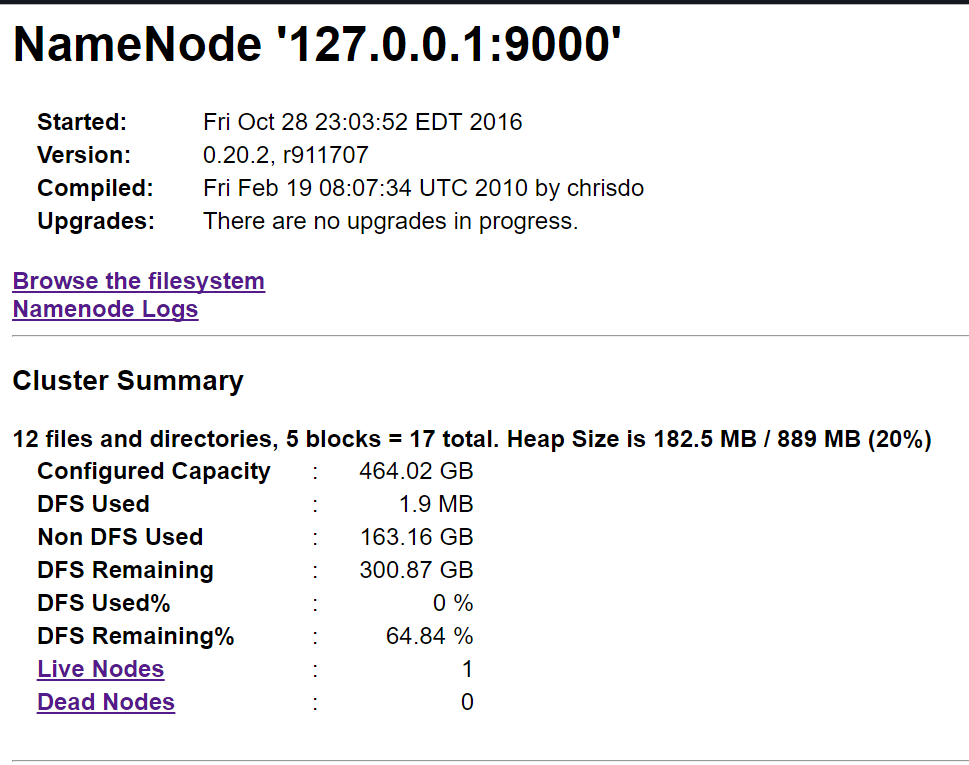


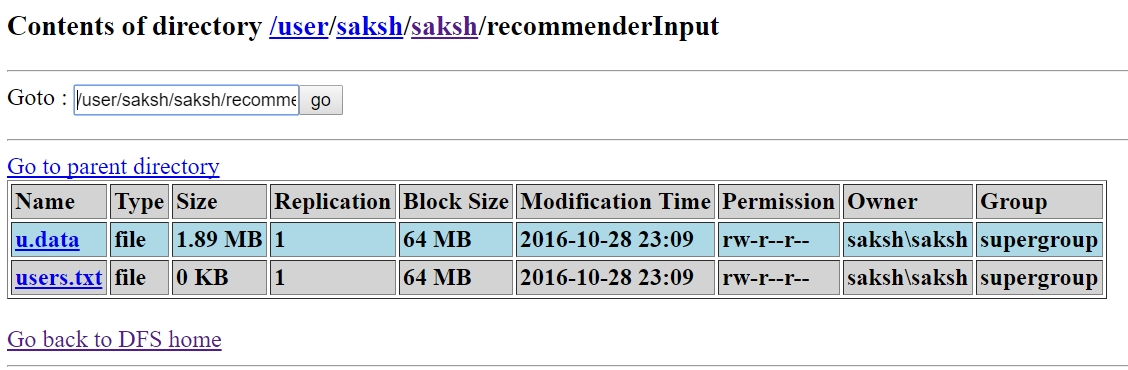


Localhost:50070



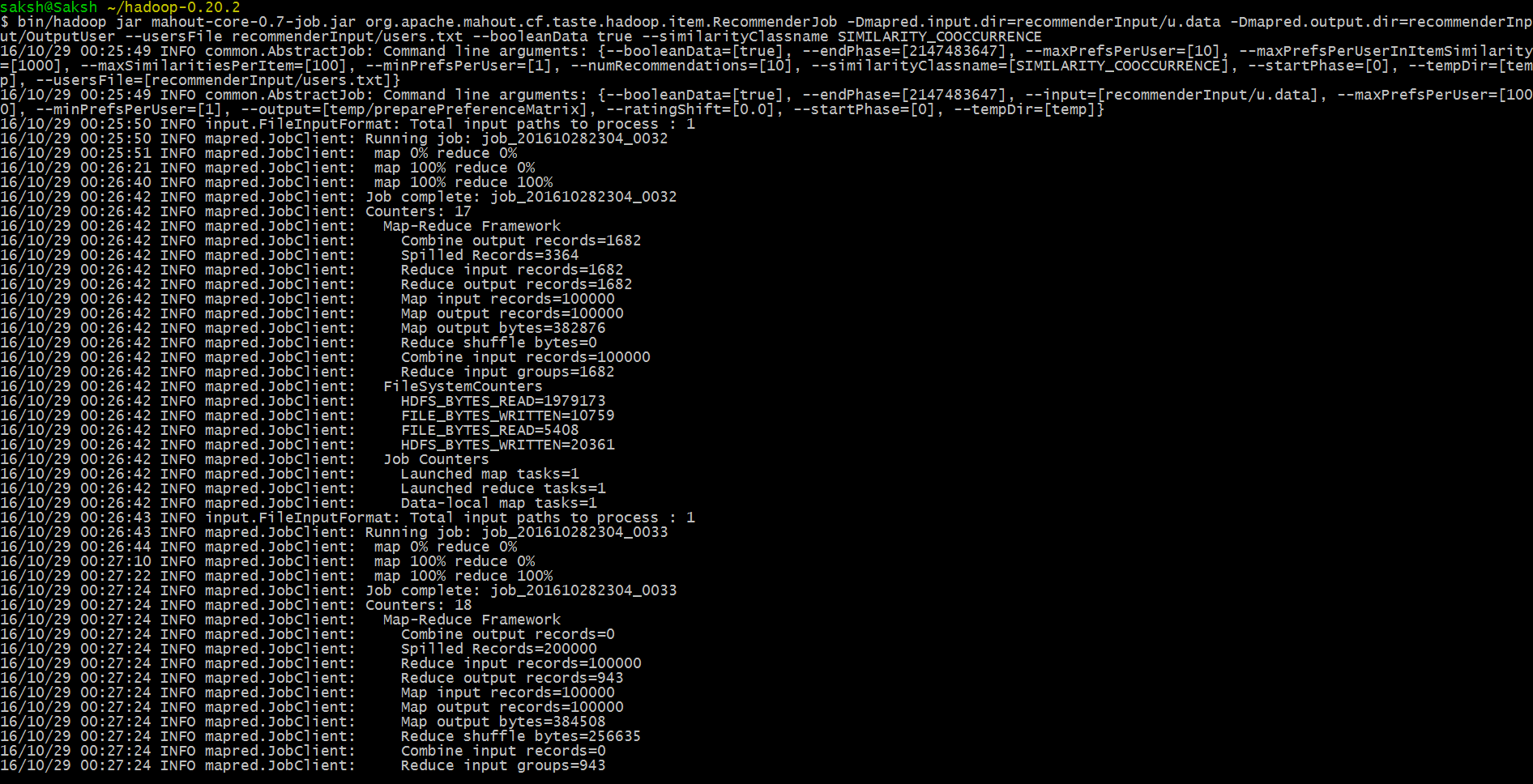
1. **Localhost:50070**

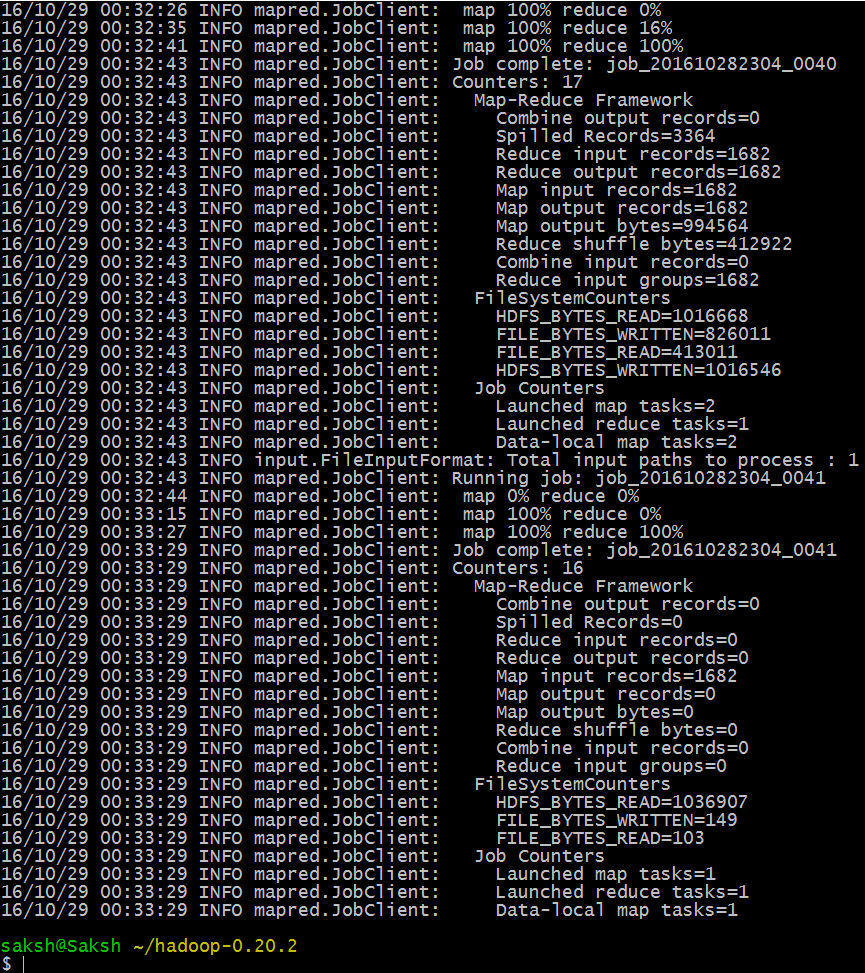




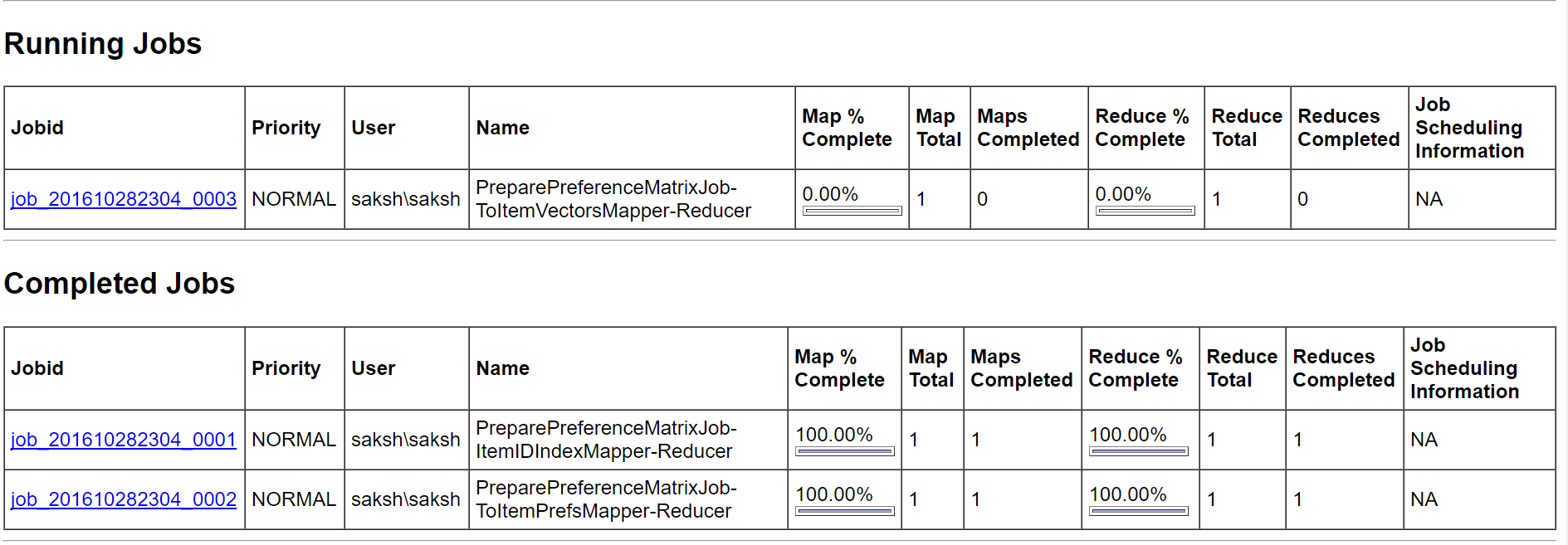
**After running the command**

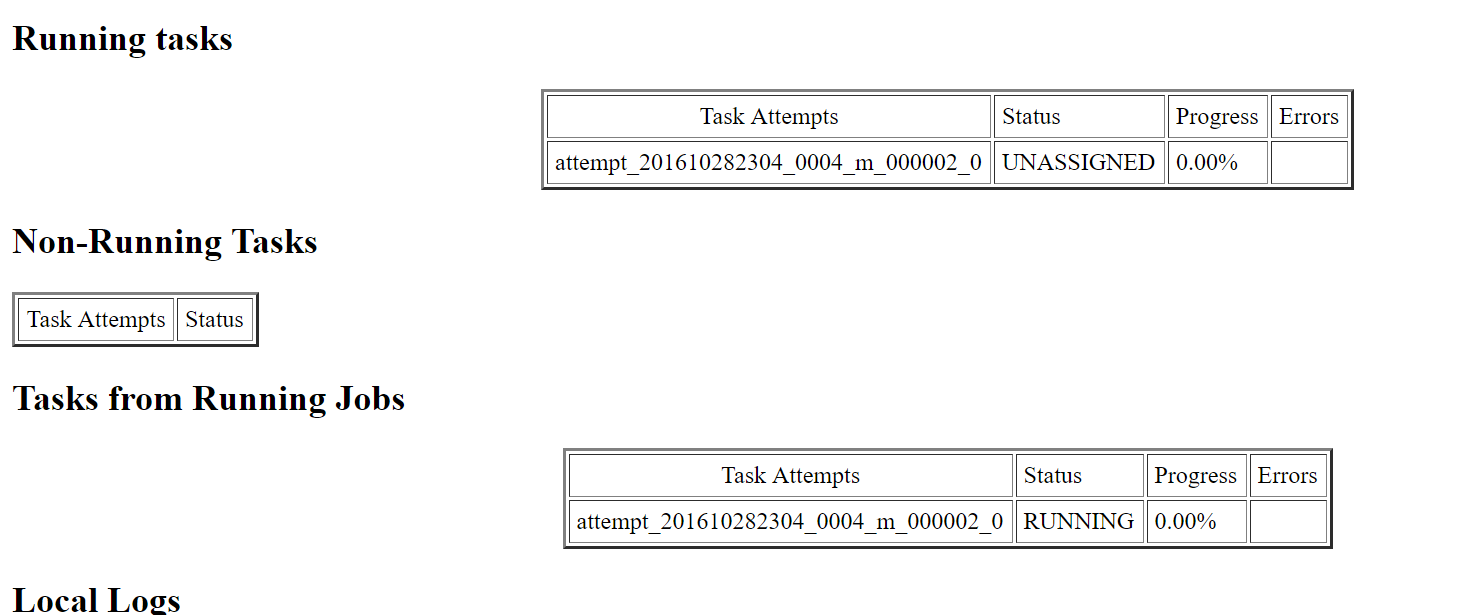
1. **Console**

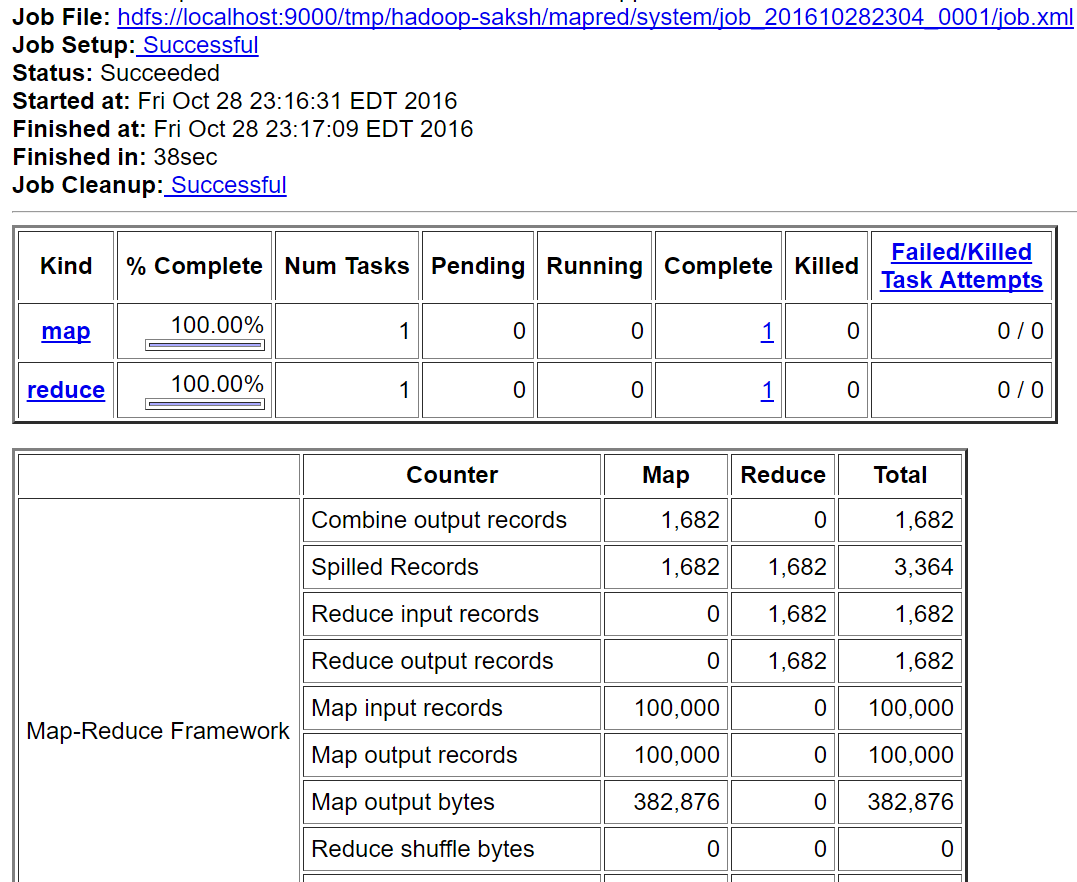




1. **Localhost:50030**

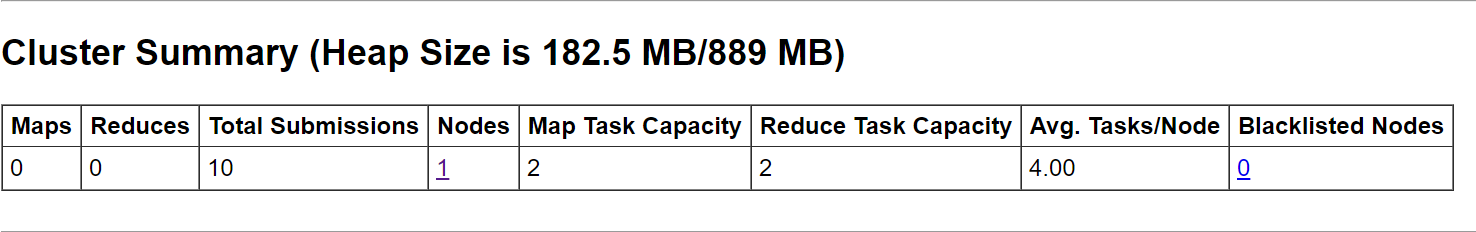


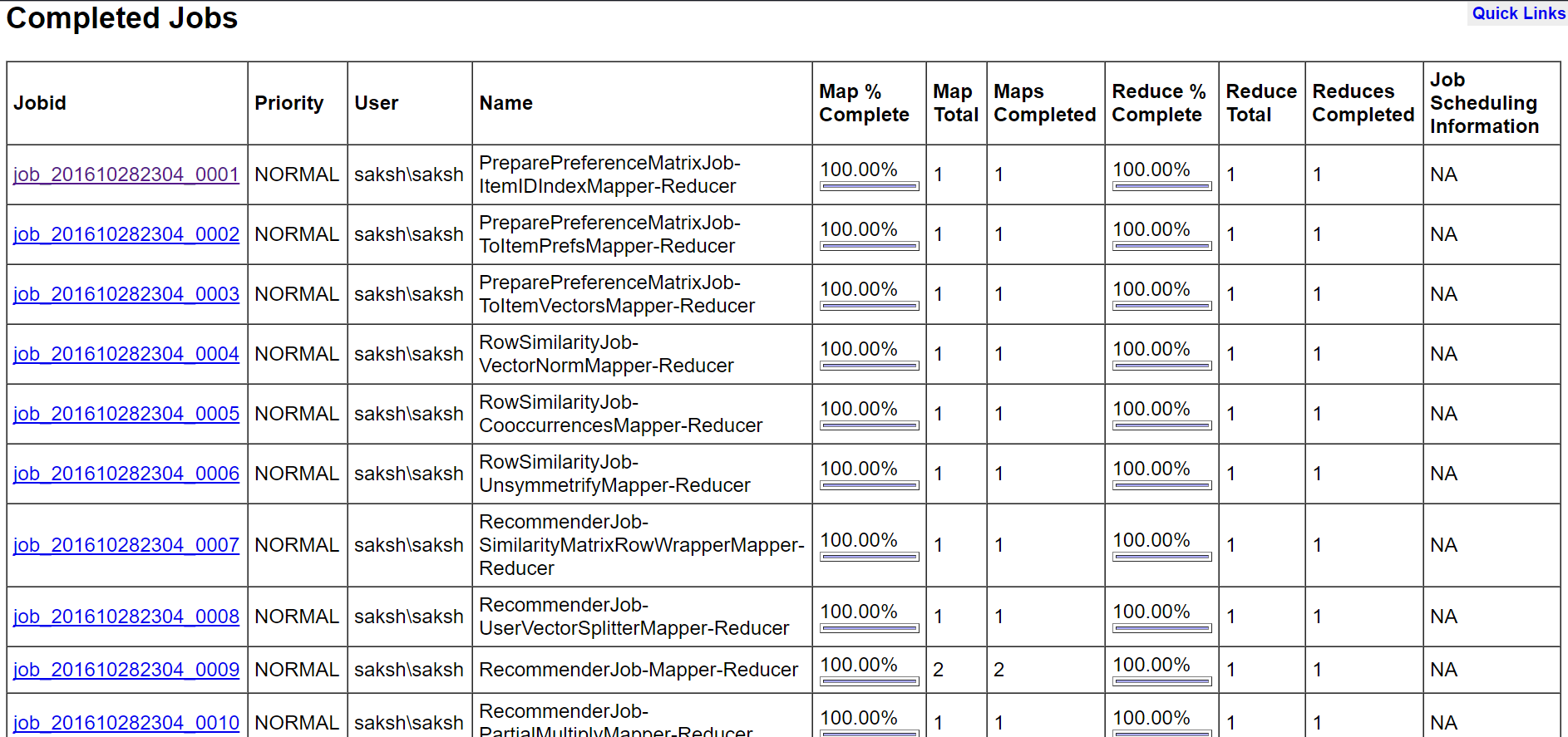




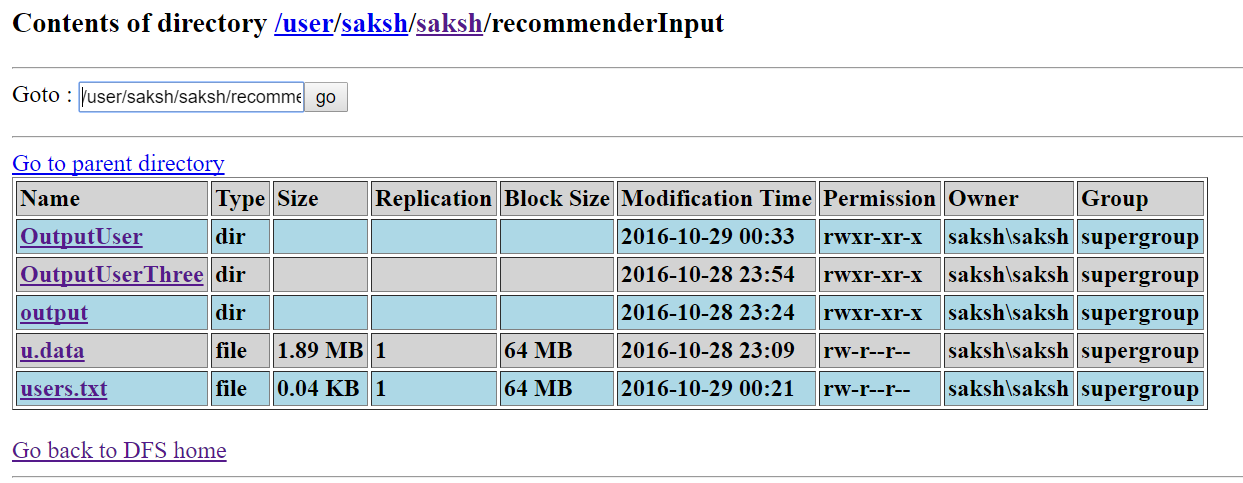


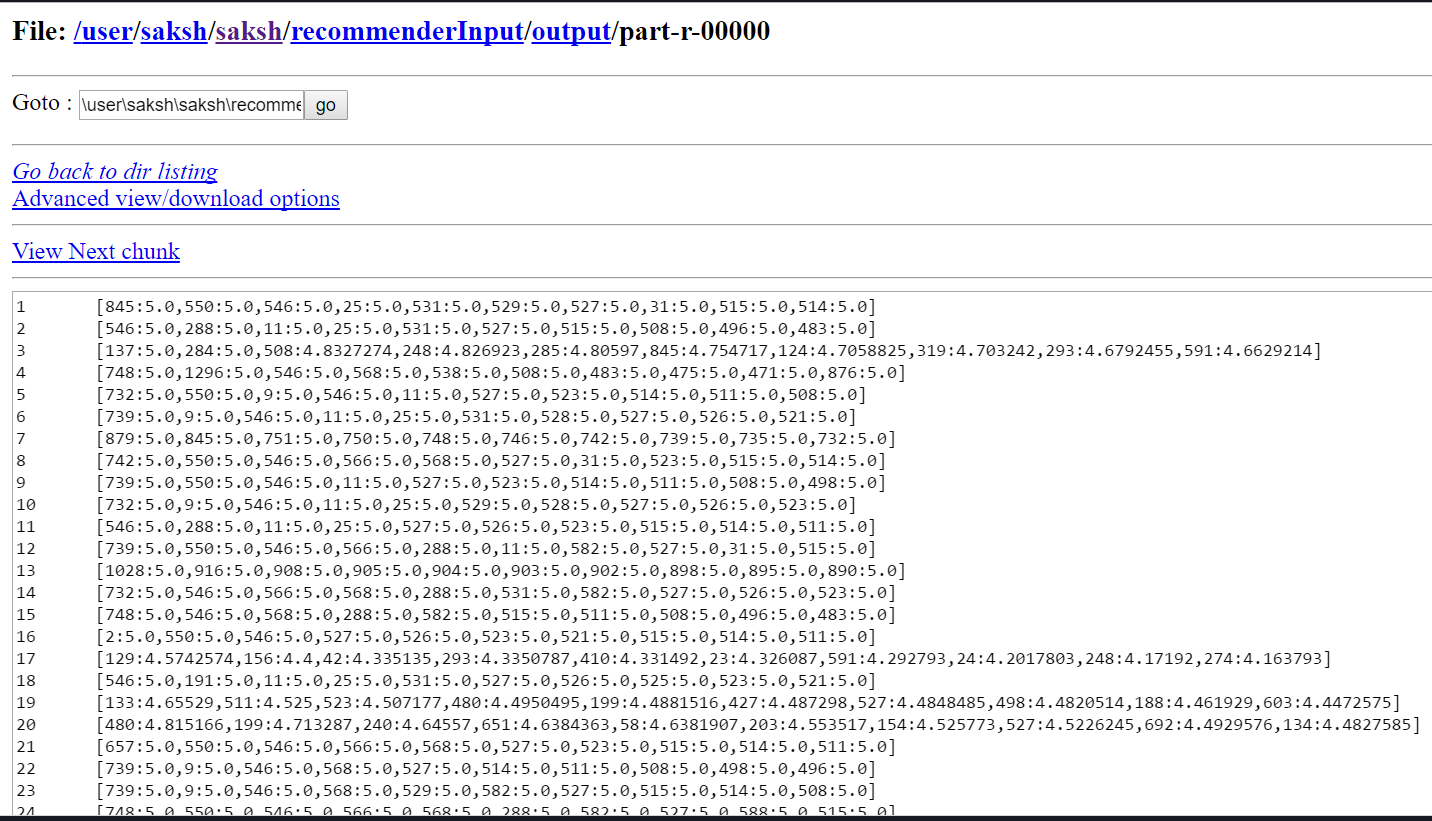
Completed Jobs





1. **Localhost:50070**

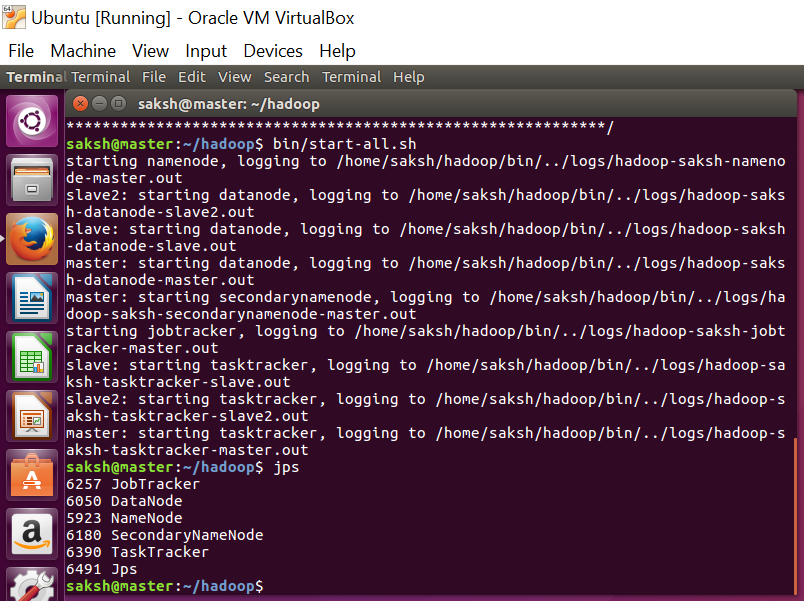


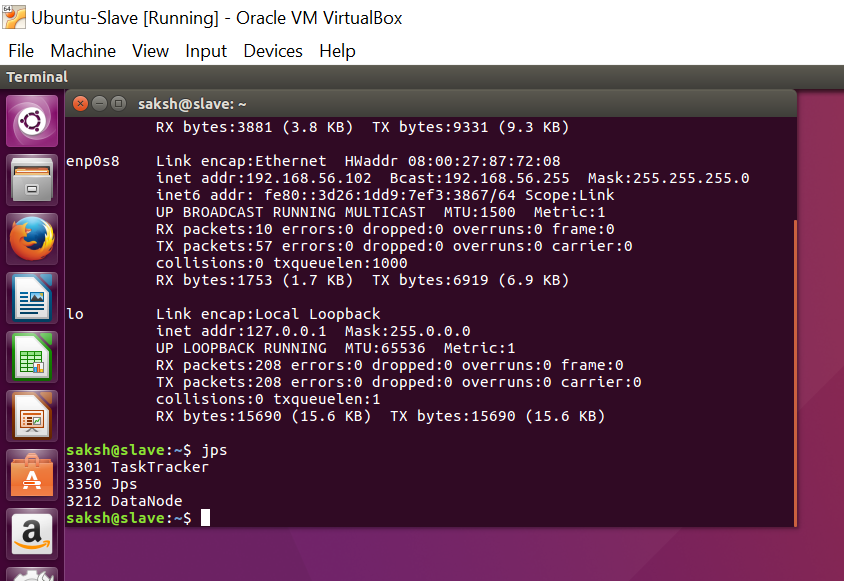




**Fully Distributed Mode Execution Snapshots**

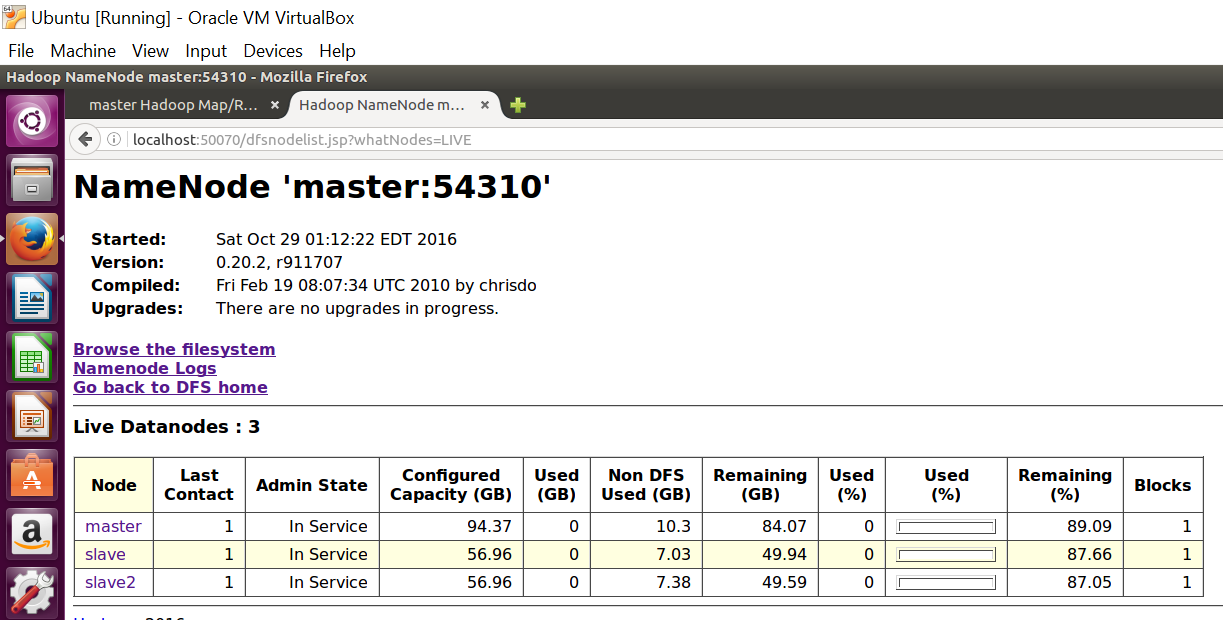
**Intial Setup**

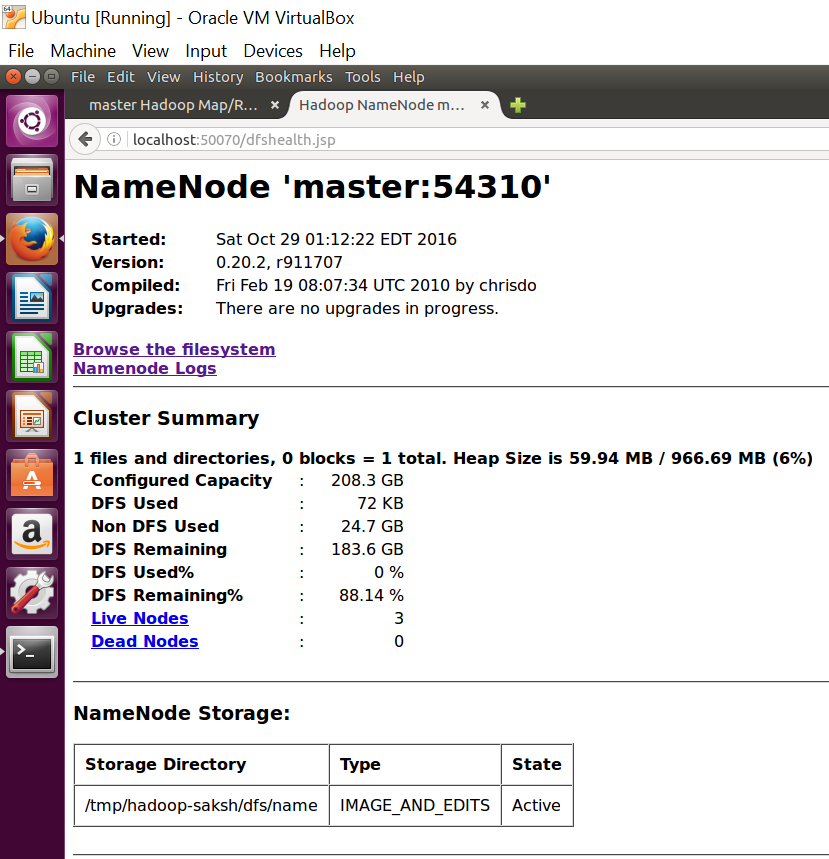




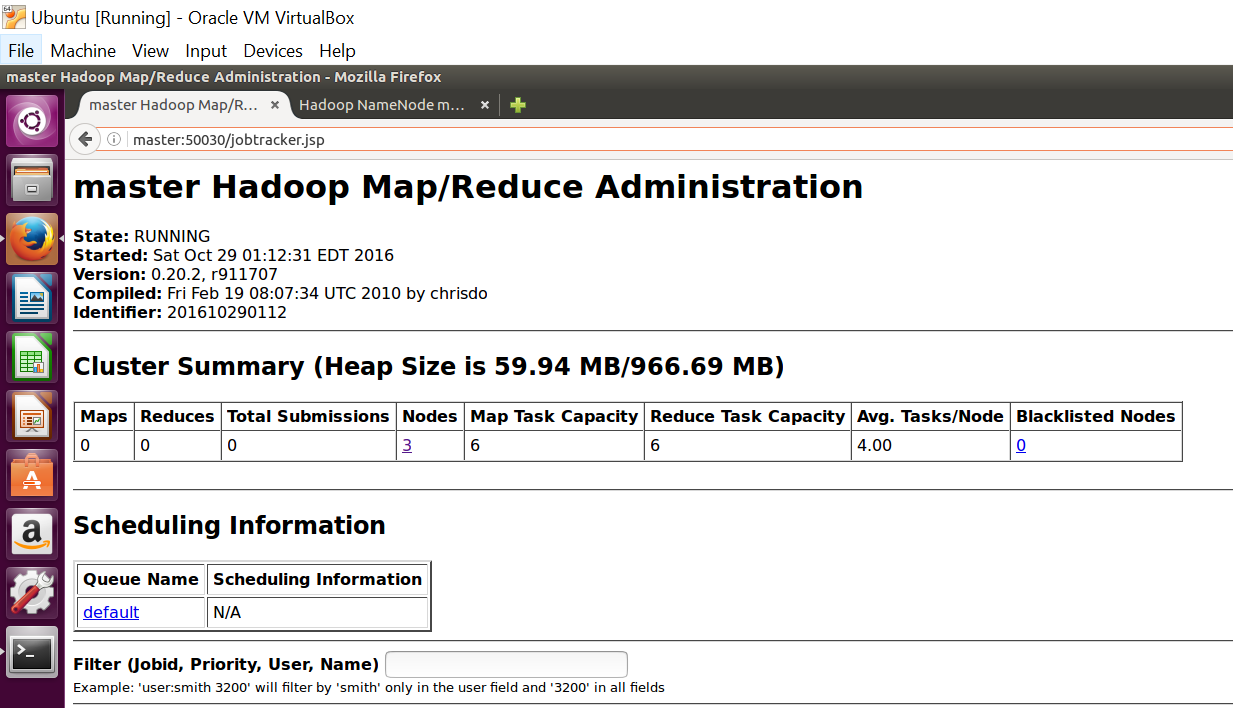


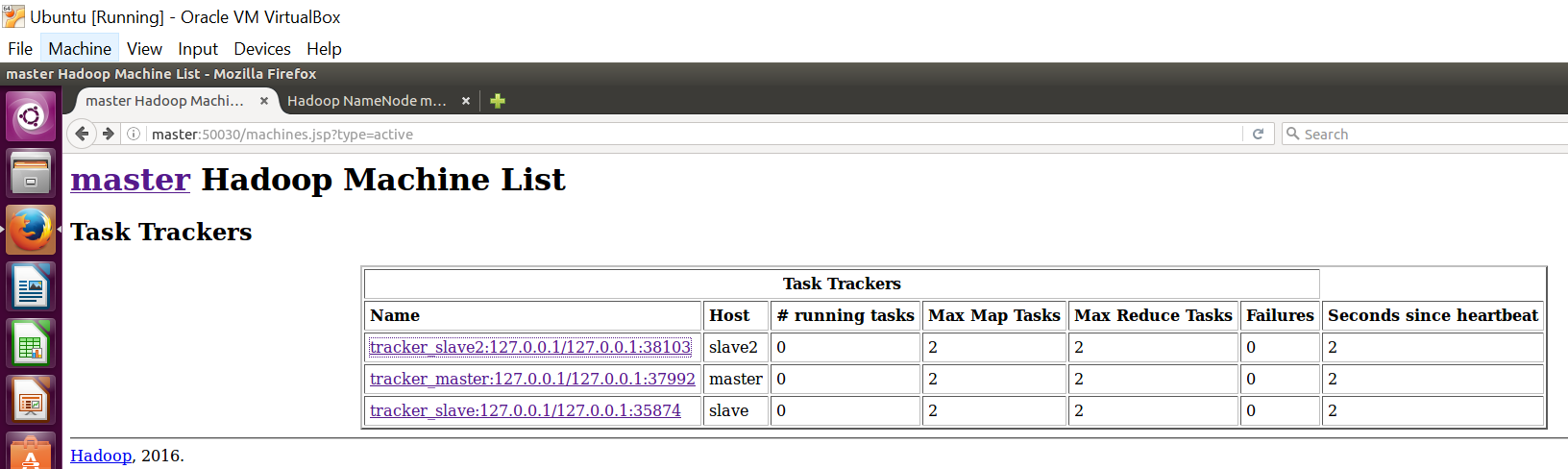
**Localhost 50070**

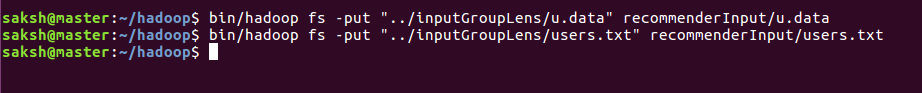


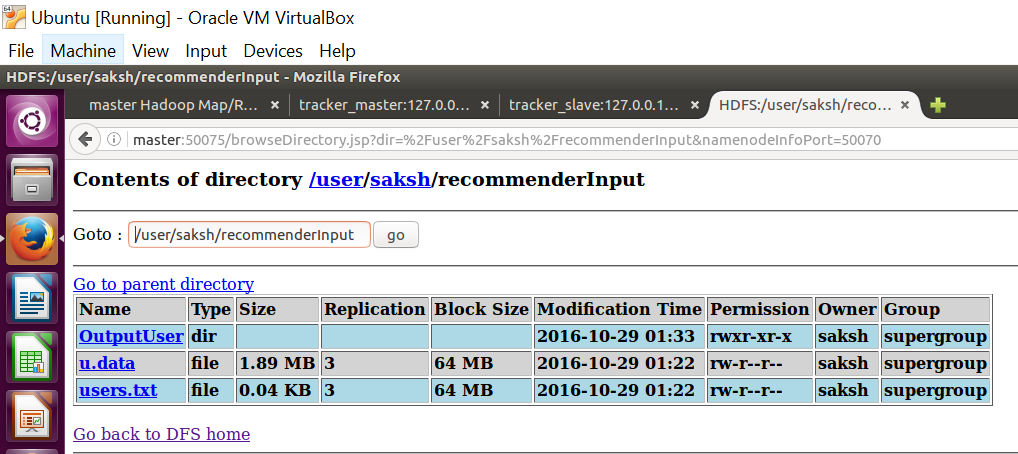


**Localhost:50030**

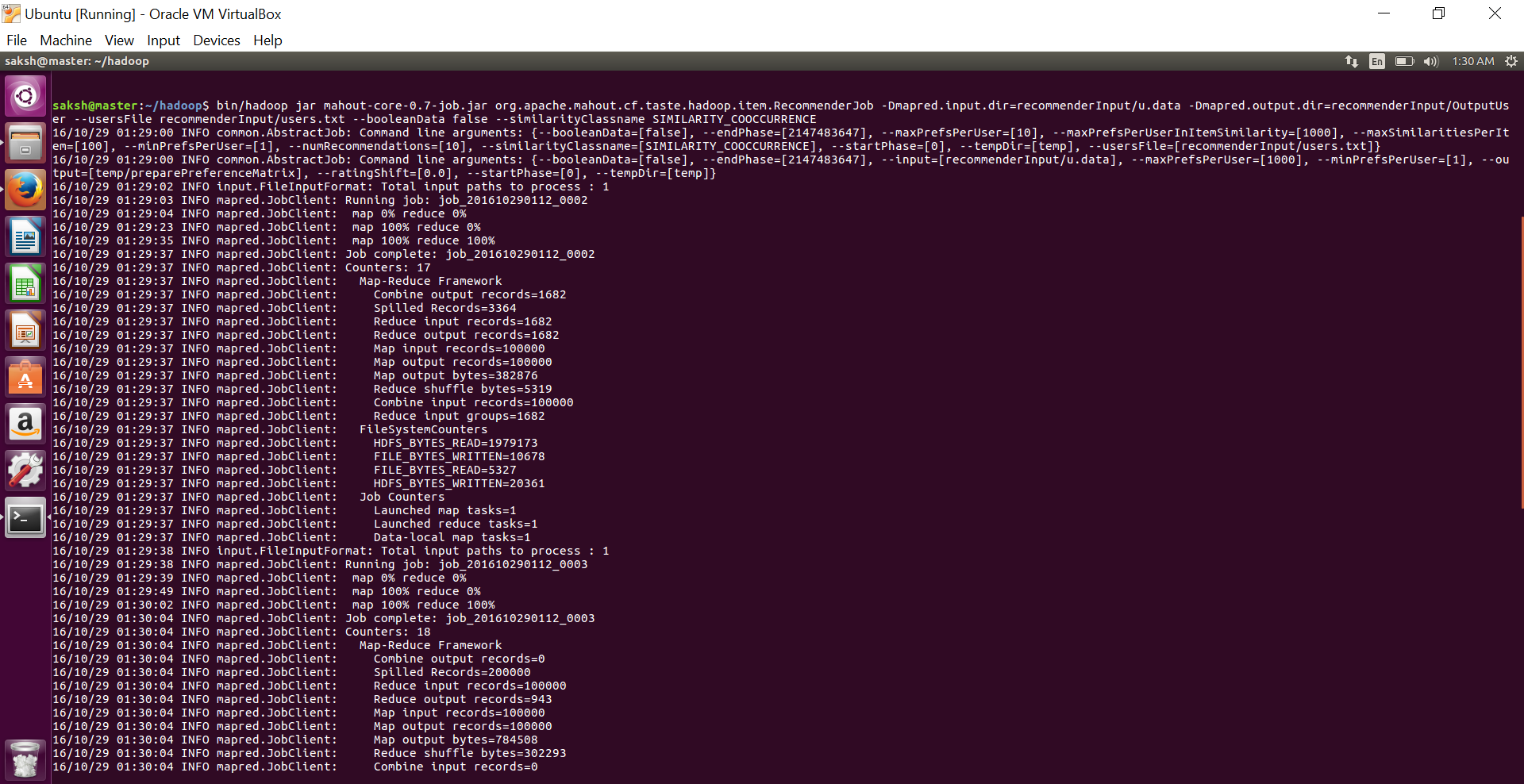


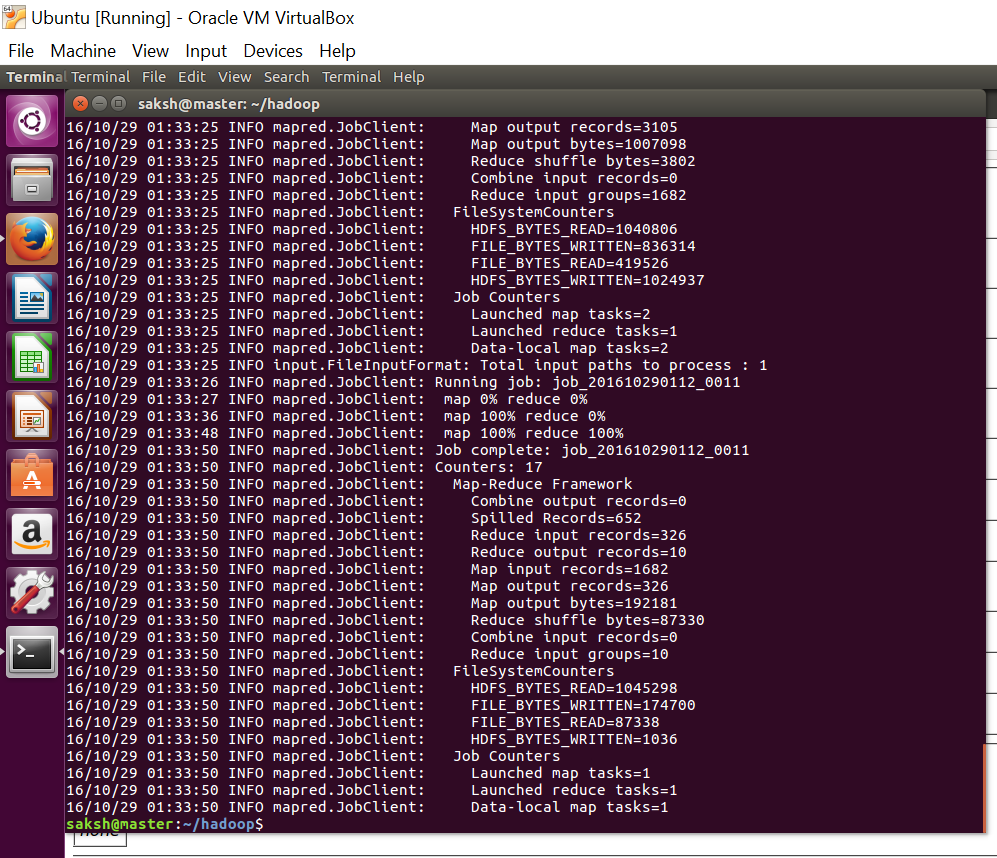


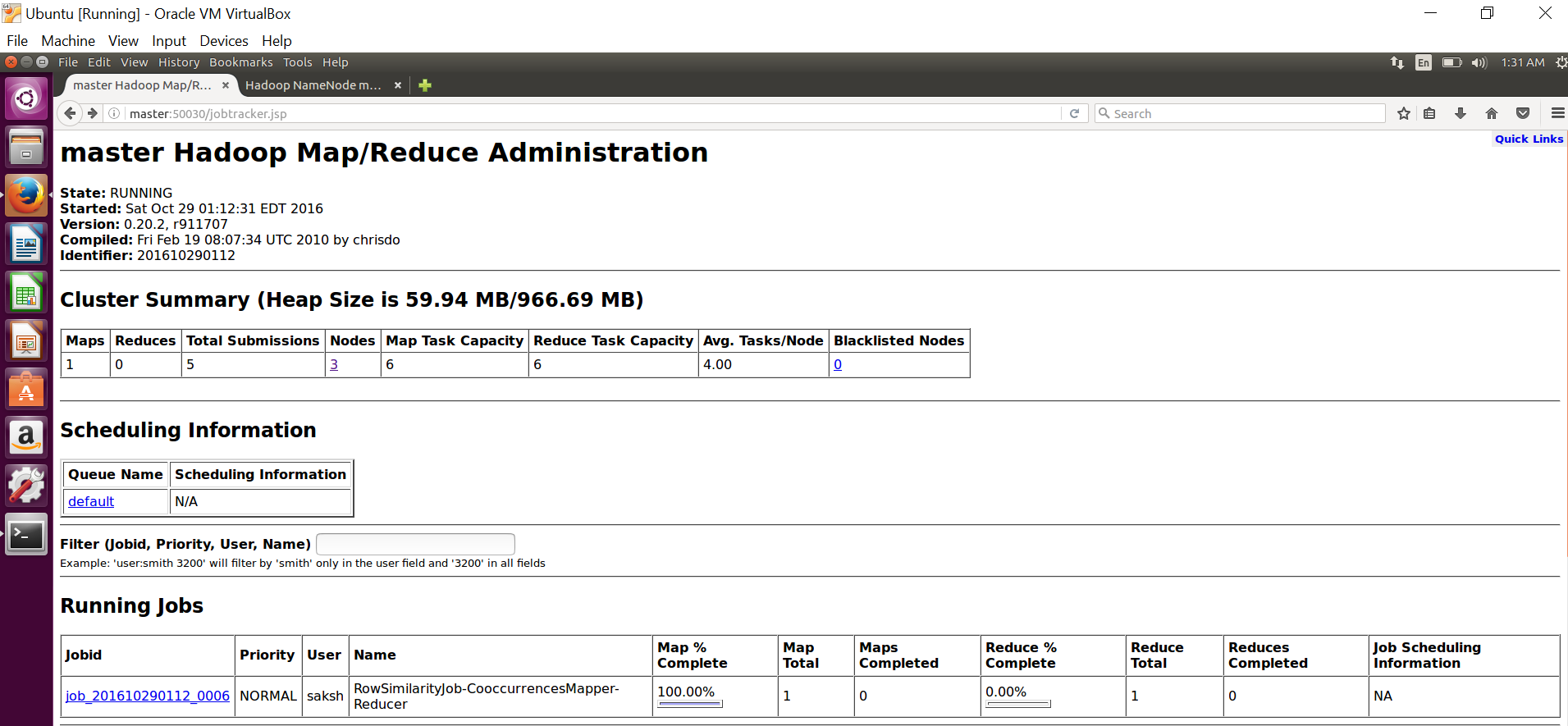


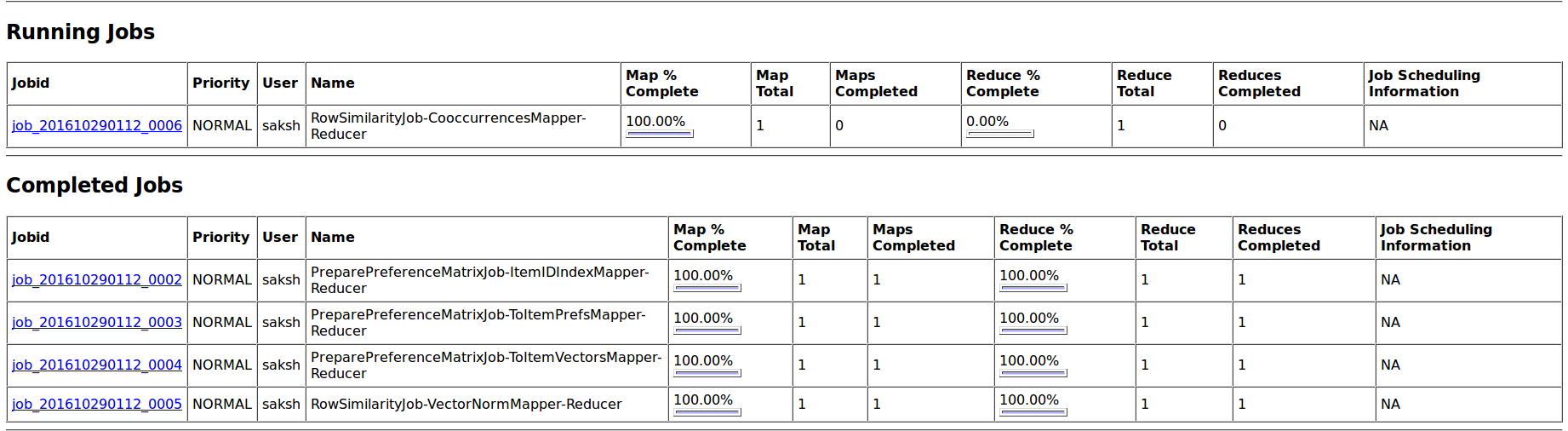


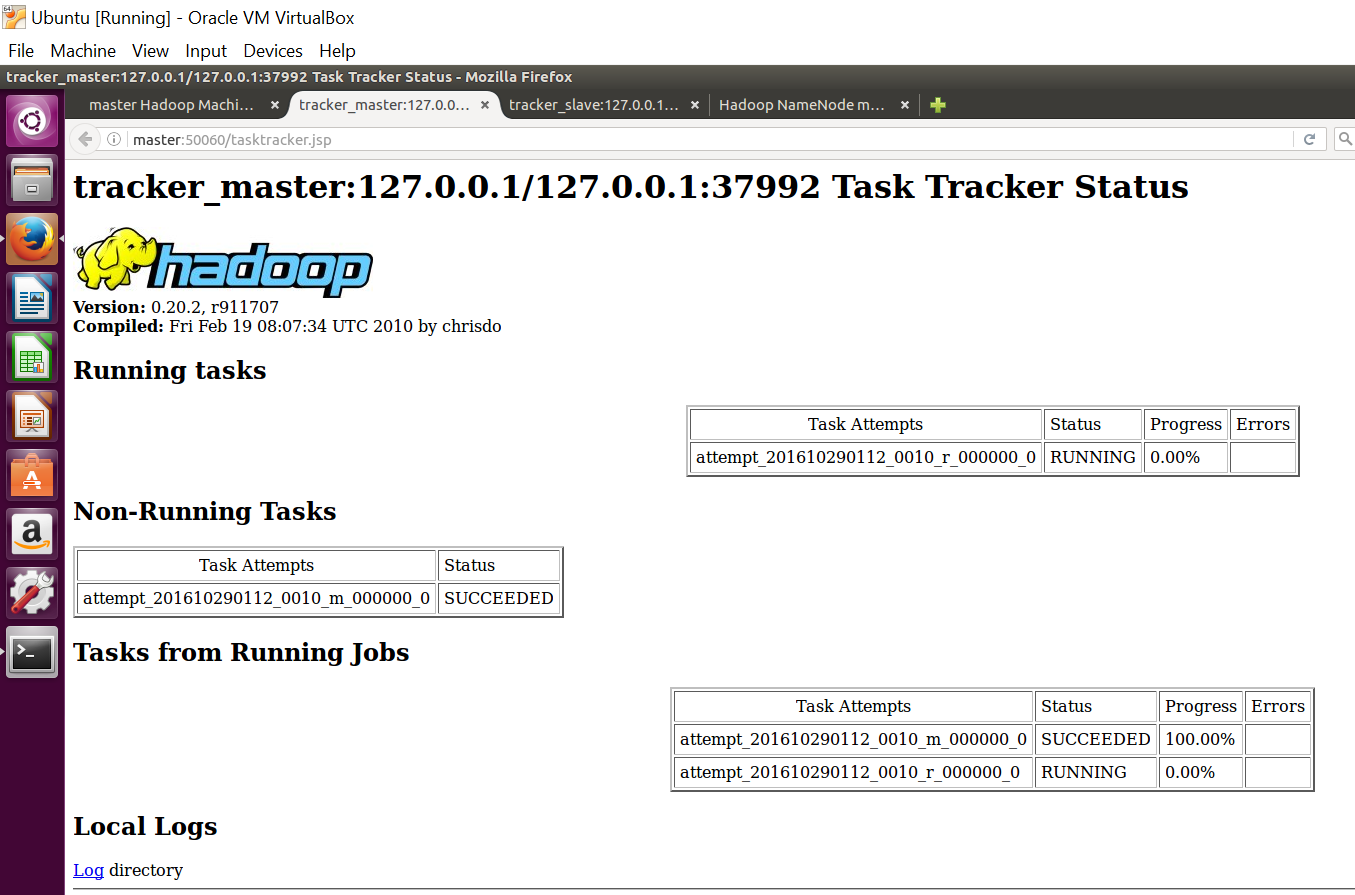
**After running**

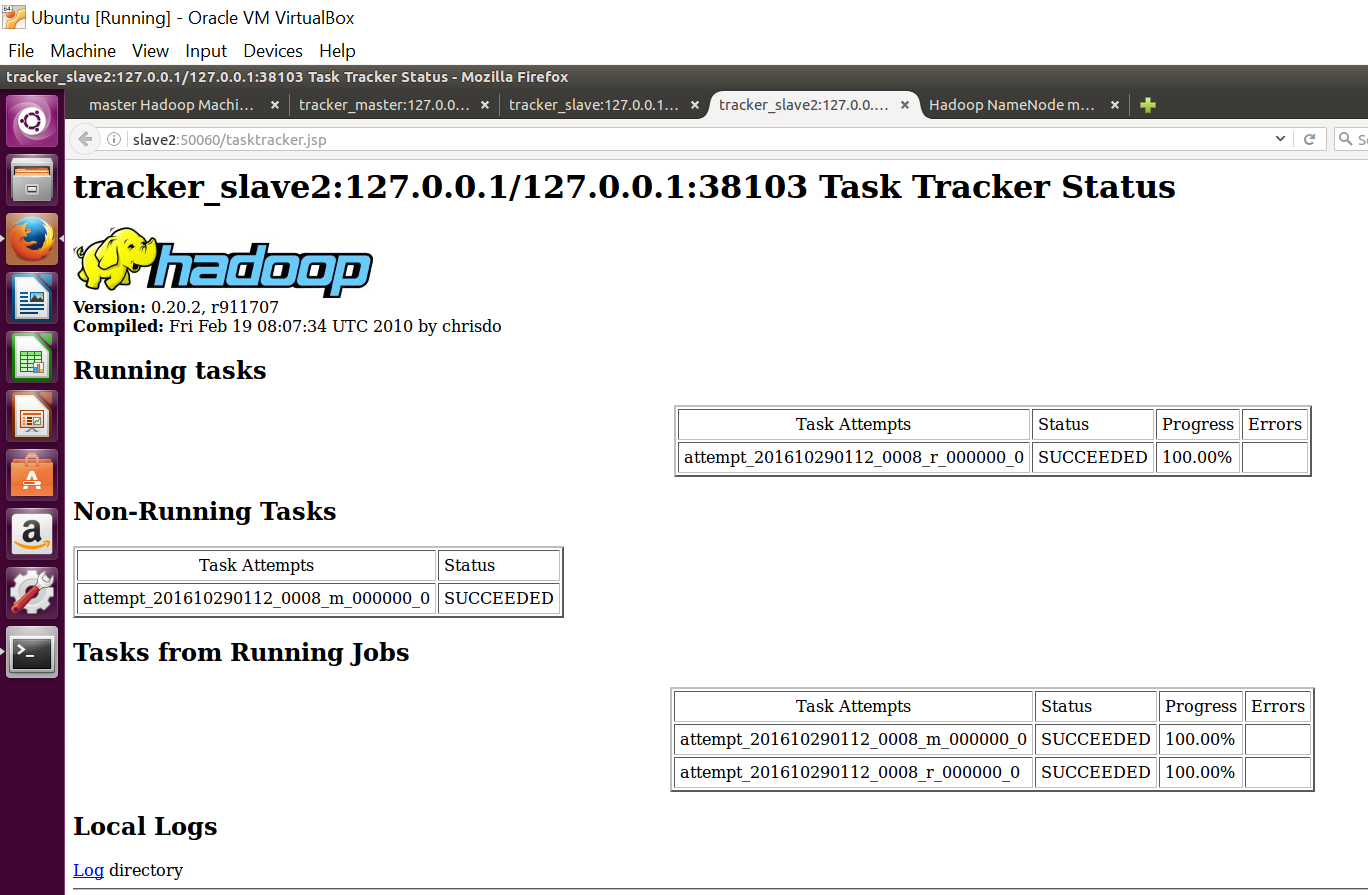




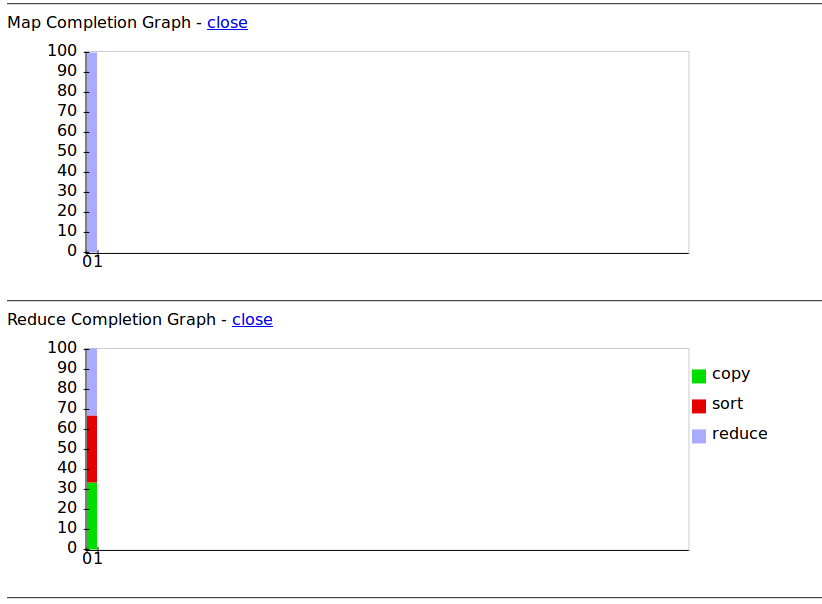




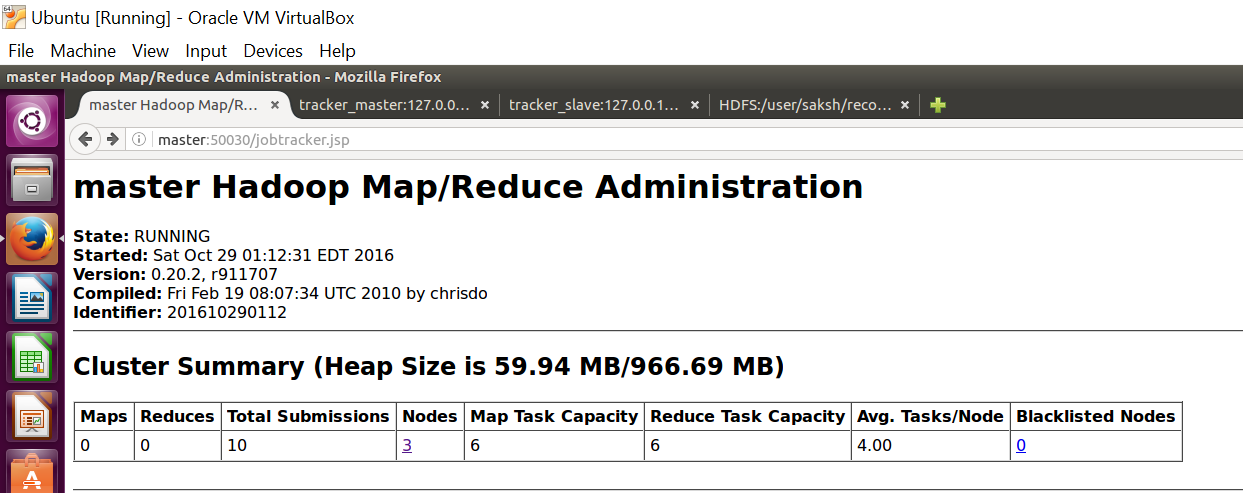


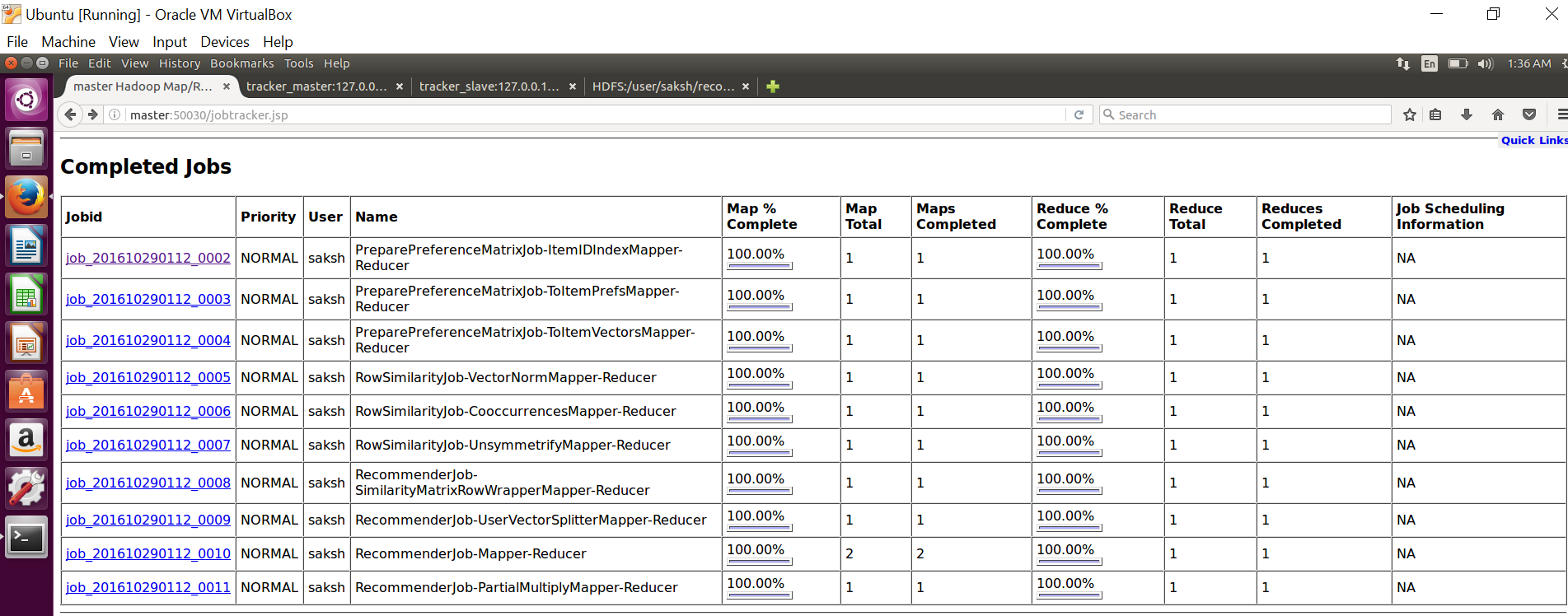


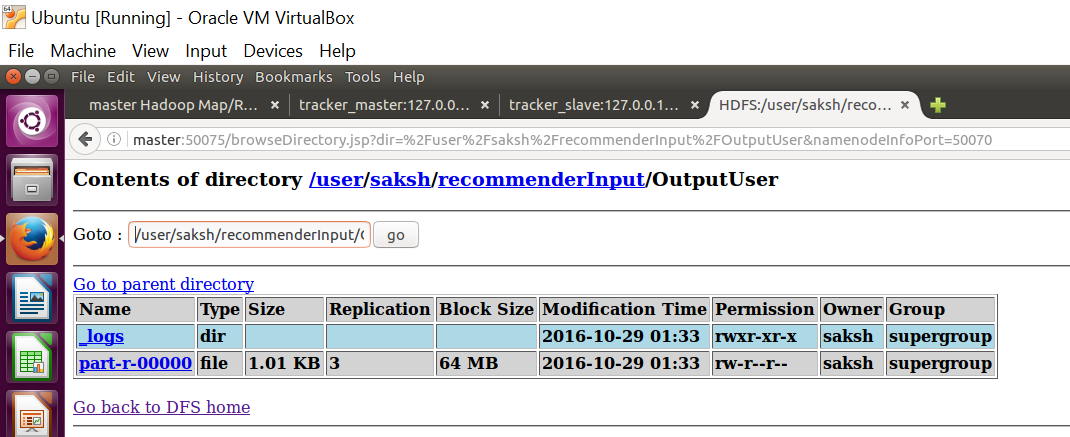


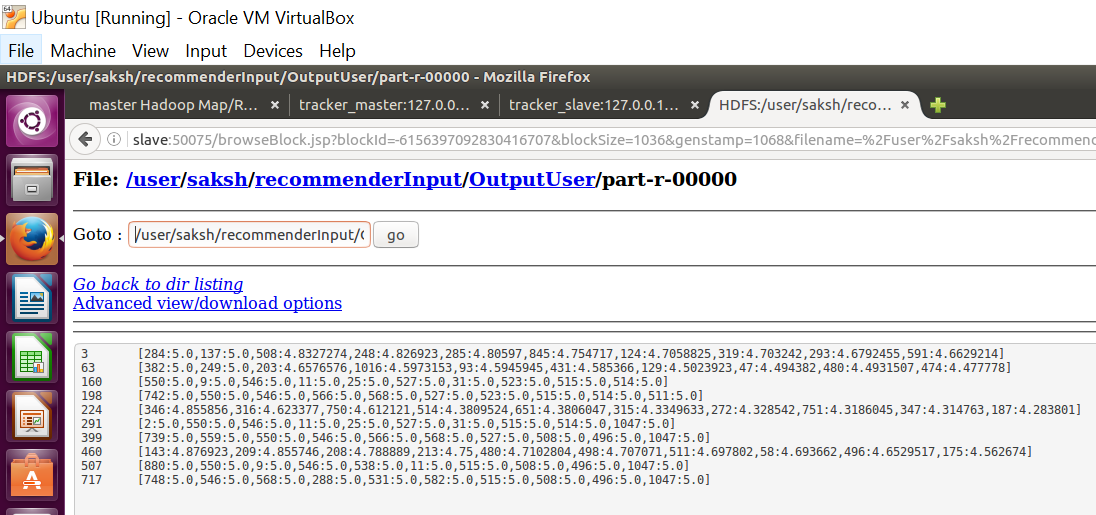


Completed Jobs

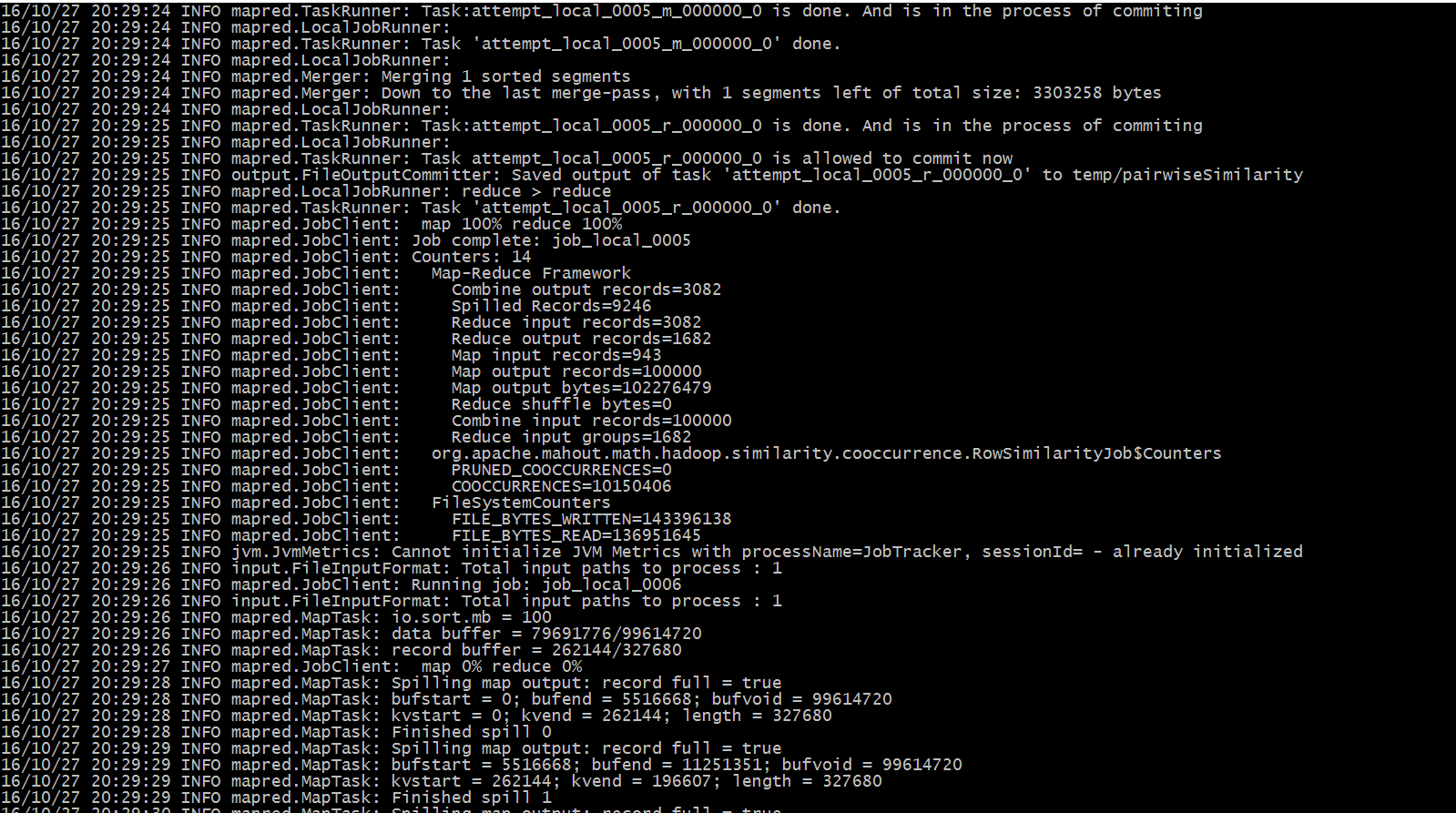


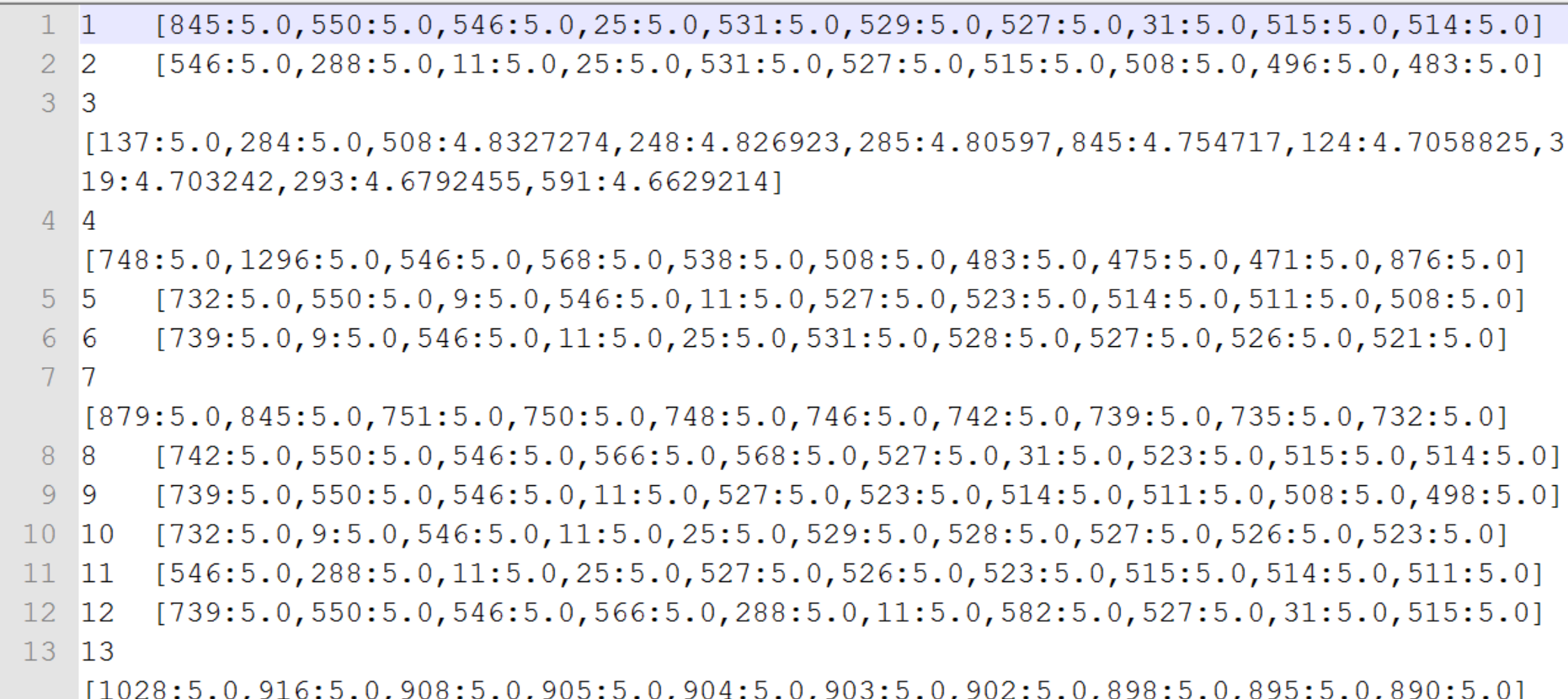


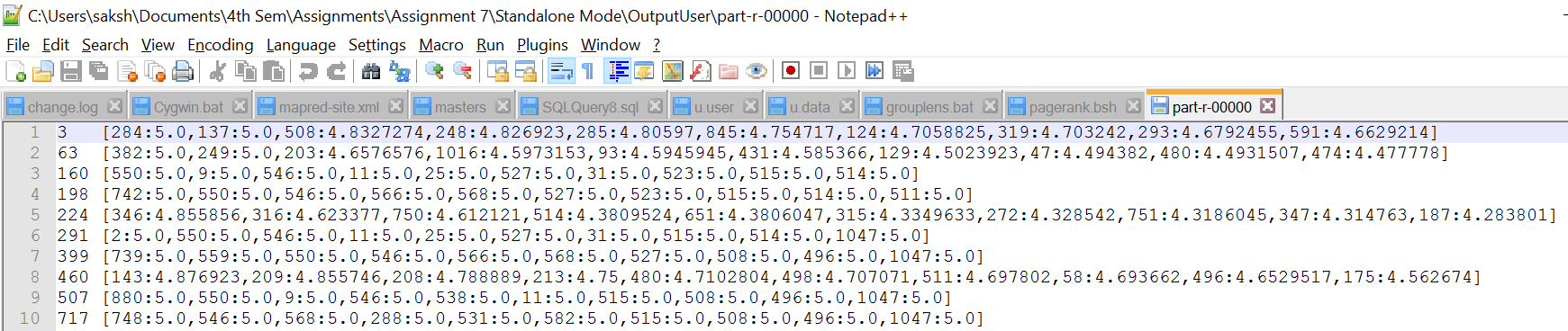




**Standalone Mode**







The python program uses the file u.data for the list of rated movies, the file u.item to get the movie titles and output.txt to get the list of recommended movies for the user.

**Conclusion**

1. **u.data** file contains all the user id, item id, rating and timestamp
2. **u.item** file contains the movie id, movie name, release year and imdb link
3. **users.txt** file contains user for which recommendation needs to be done
4. **output** directory contains the recommendation for all users
5. **OutputUser** directory contain recommendation for selected users, which are given in users.txt file