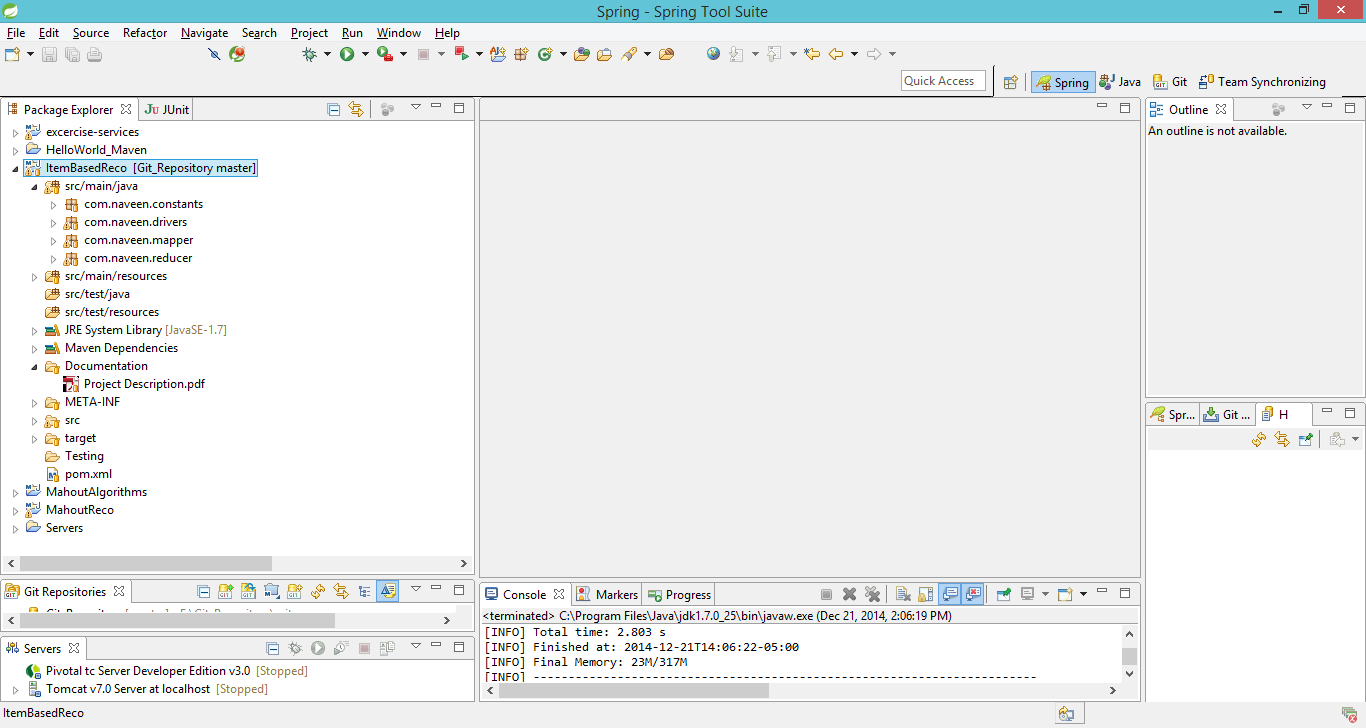
**Testing Procedure**

1. The code has been developed using Apache Maven and Spring Test Suite. Any integrated development environment can be used for generating a JAR file.
2. Checkout the Folder(ItemBasedReco) from the GitHUb location into a JAVA project in any of the IDE

<https://github.com/cloud-class-projects/recommendation>

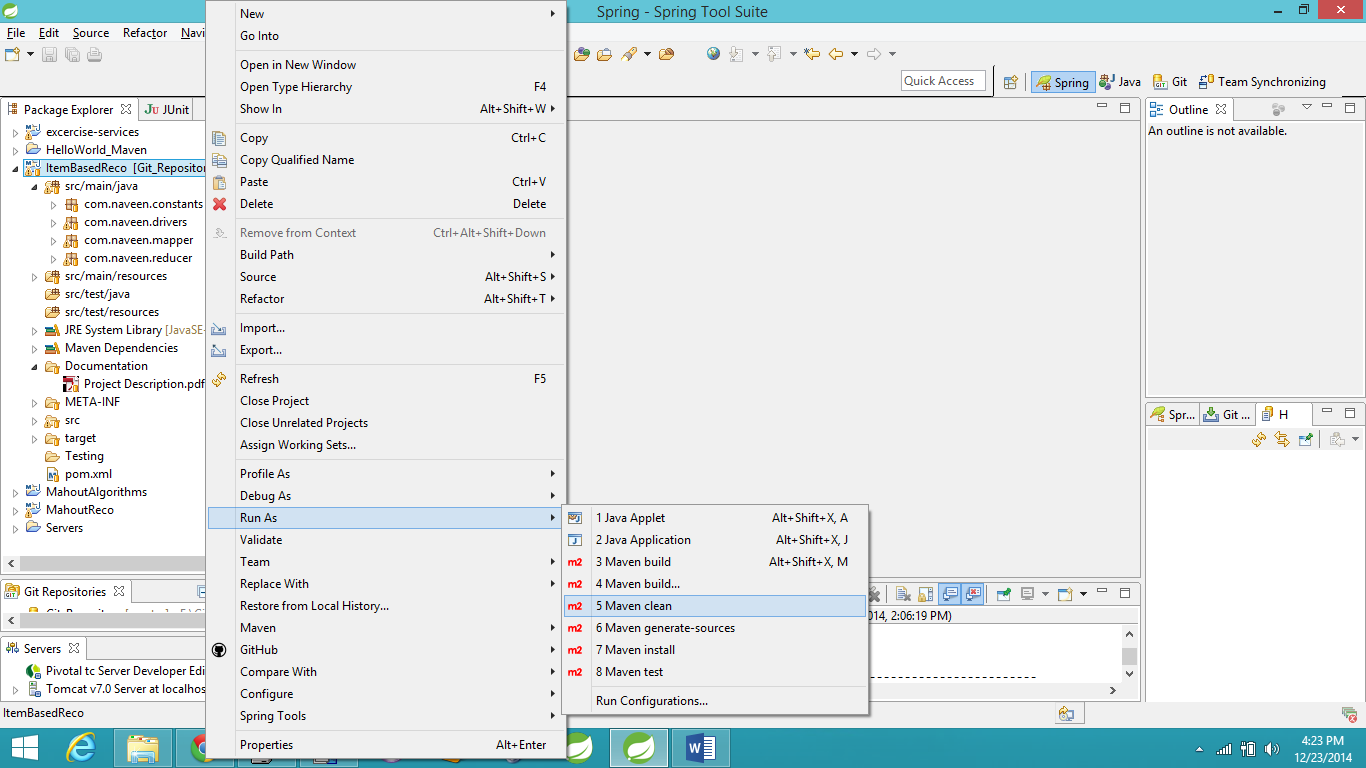
As in below screenshot of Spring Test Suite,

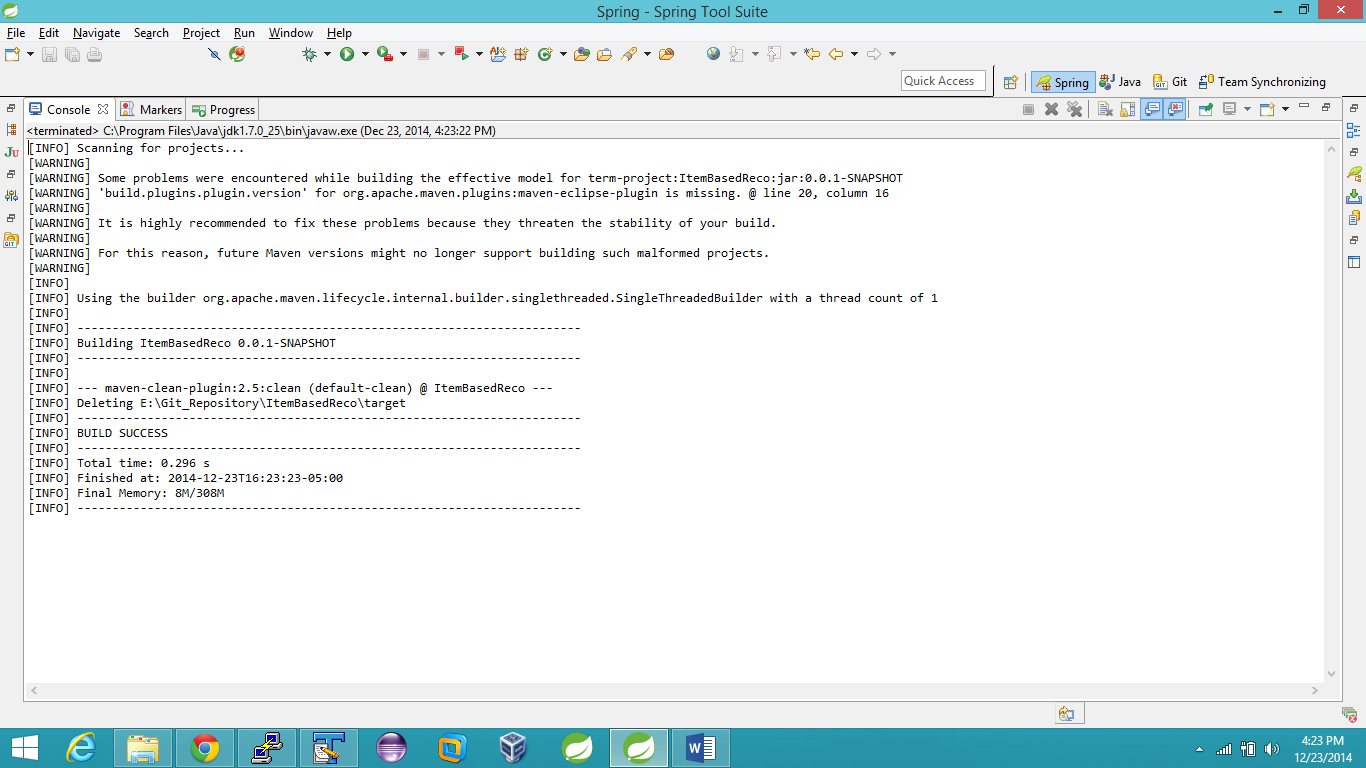


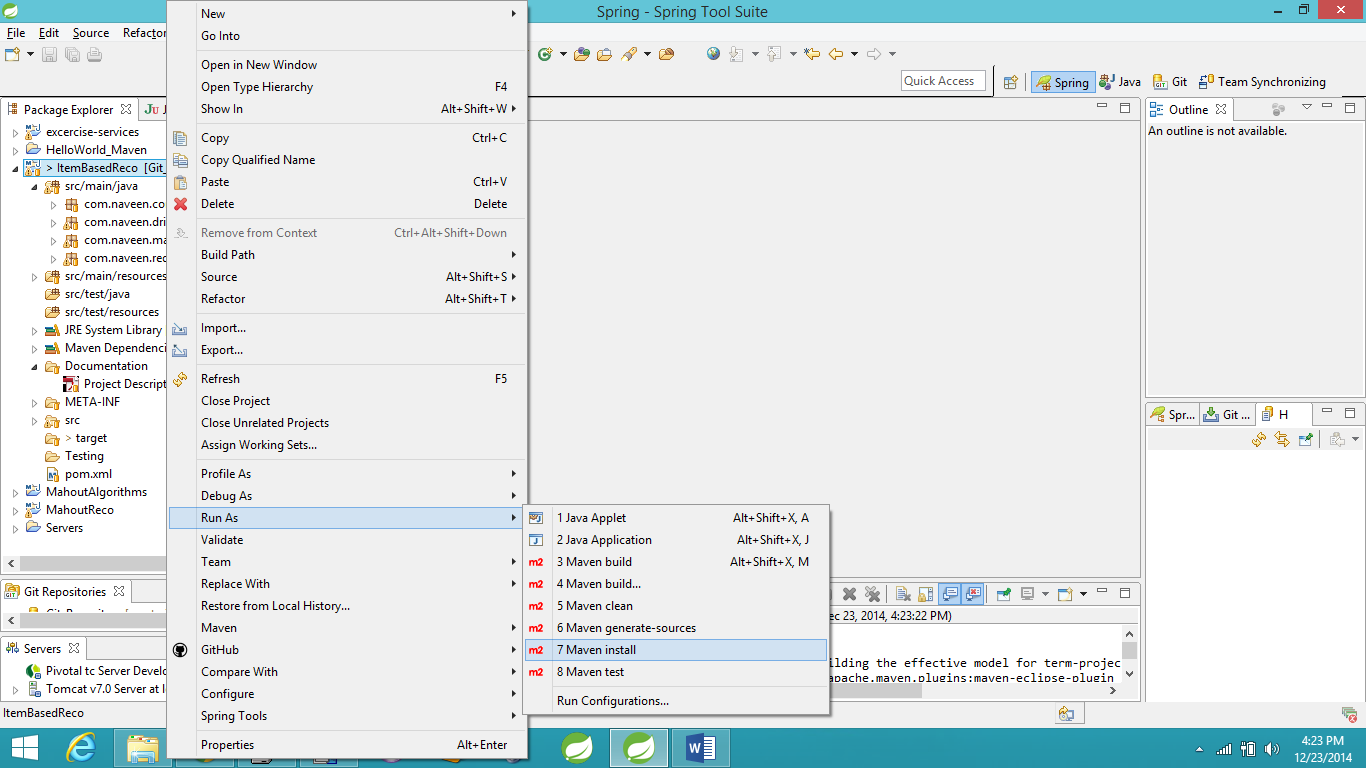
We need to generate JAR file out of the Maven Project. Right click on the project

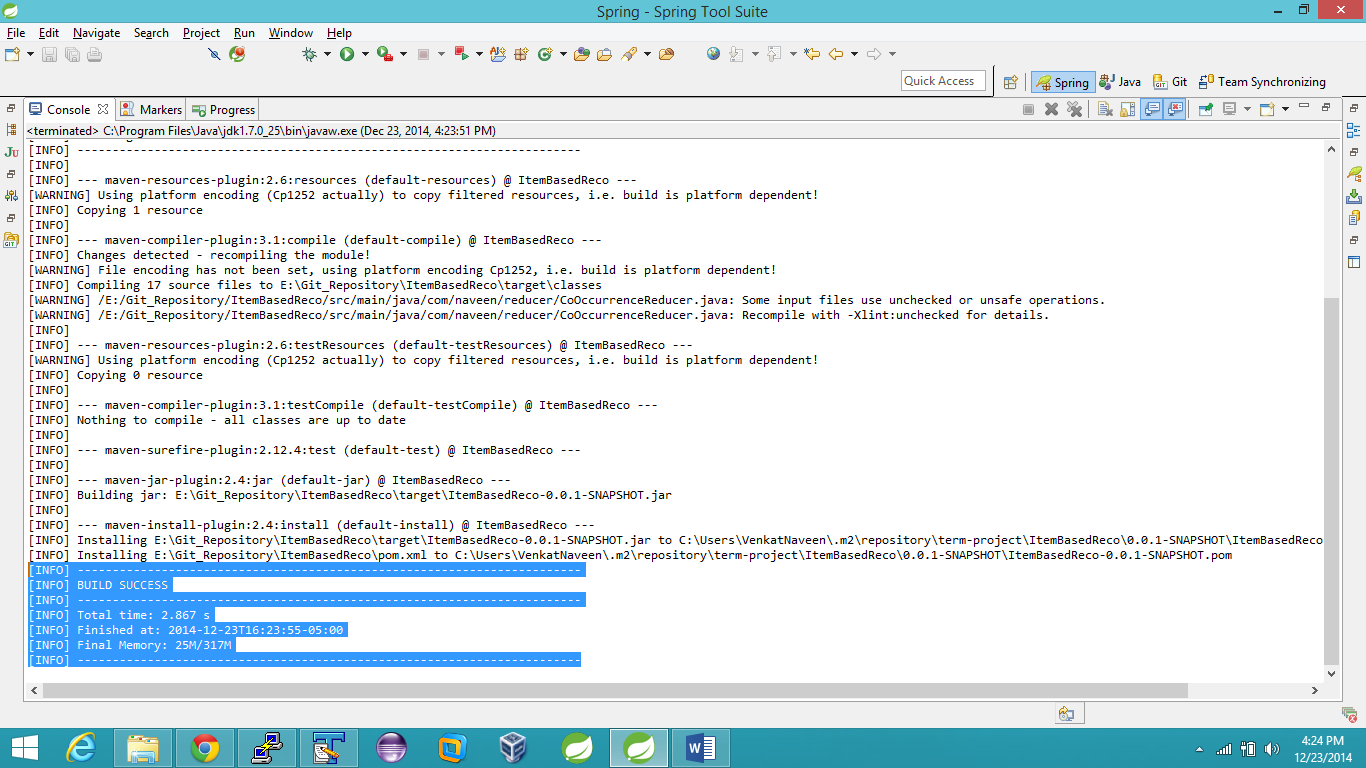
Run As -> Maven Clean

Run As -> Maven Install









The above BUILD Success message should be displayed to confirm the creation of JAR file.

The JAR is located under the target folder.

Already generated JAR is present in the GIT HUB location

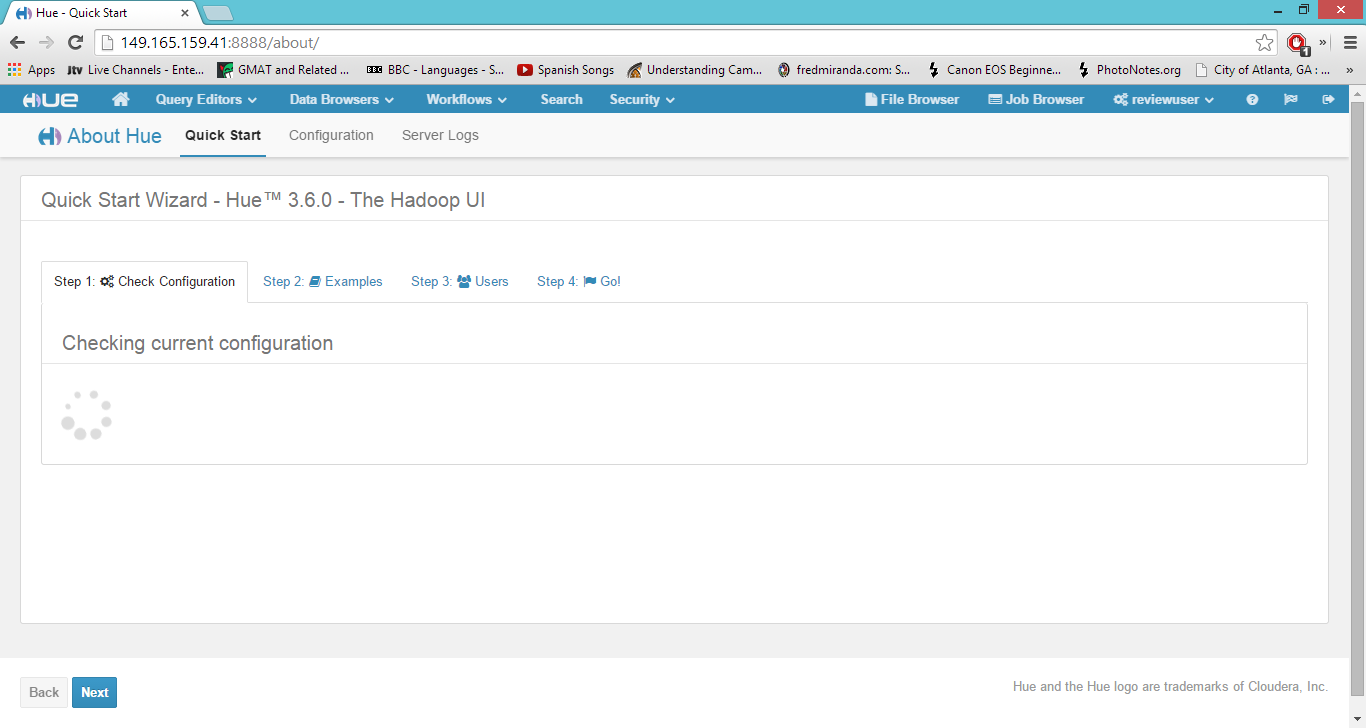
<https://github.com/cloud-class-projects/recommendation/tree/master/ItemBasedReco/target>

1. Hue (Web based application) should be used for submitting the Oozie jobs.

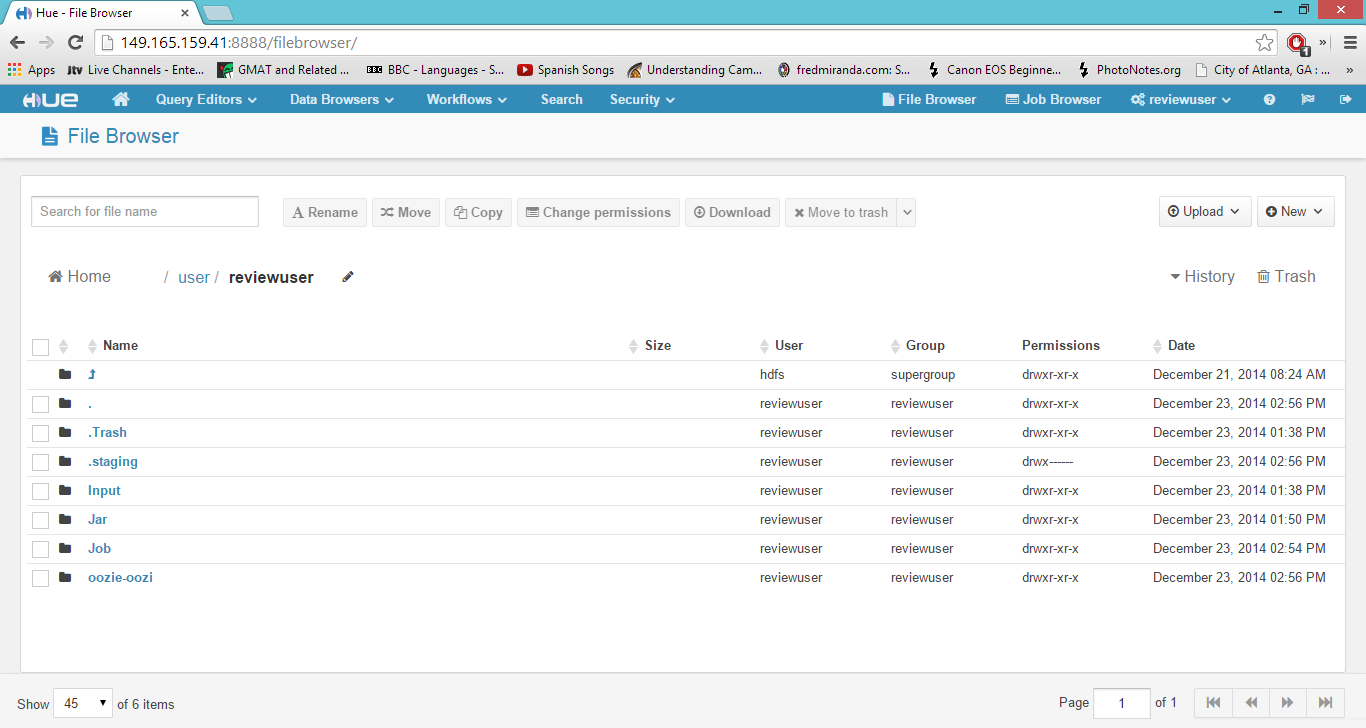
<http://149.165.159.41:8888/accounts/login/?next=/>

User – reviewuser

Password – reviewuser



“File Browser” tab should be used for accessing HDFS cluster.



The following folders should be created for the jobs to run,

/Input for storing the input files

/Jar for storing the JAR file after downloading from <https://github.com/cloud-class-projects/recommendation/tree/master/ItemBasedReco/target>

1. The below attached shell script should be executed for all the map reduce jobs to be executed.

Login to the command line of any one of the data node,

Execute the shell script,



Contents of the above shell script,

#!/bin/bash

sudo -u hdfs hadoop fs -rm -R /user/hdfs/jar

sudo -u hdfs hadoop fs -rm -R /user/hdfs/Input

sudo -u hdfs hadoop fs -mkdir /user/hdfs/jar

sudo -u hdfs hadoop fs -mkdir /user/hdfs/Input

sudo -u hdfs hadoop fs -cp /user/reviewuser/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar /user/hdfs/jar/

sudo -u hdfs hadoop fs -cp /user/reviewuser/Input/u1.base /user/hdfs/Input/

cd ~

rm -R ~/Jar

mkdir ~/Jar

chmod -R 777 ~/Jar

sudo -u hdfs hadoop fs -copyToLocal /user/reviewuser/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar ~/Jar/

sudo -u hdfs hadoop jar ~/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar com.naveen.drivers.DataScrub 1 2

sudo -u hdfs hadoop jar ~/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar com.naveen.drivers.CoOccurrenceMatrix 1 2

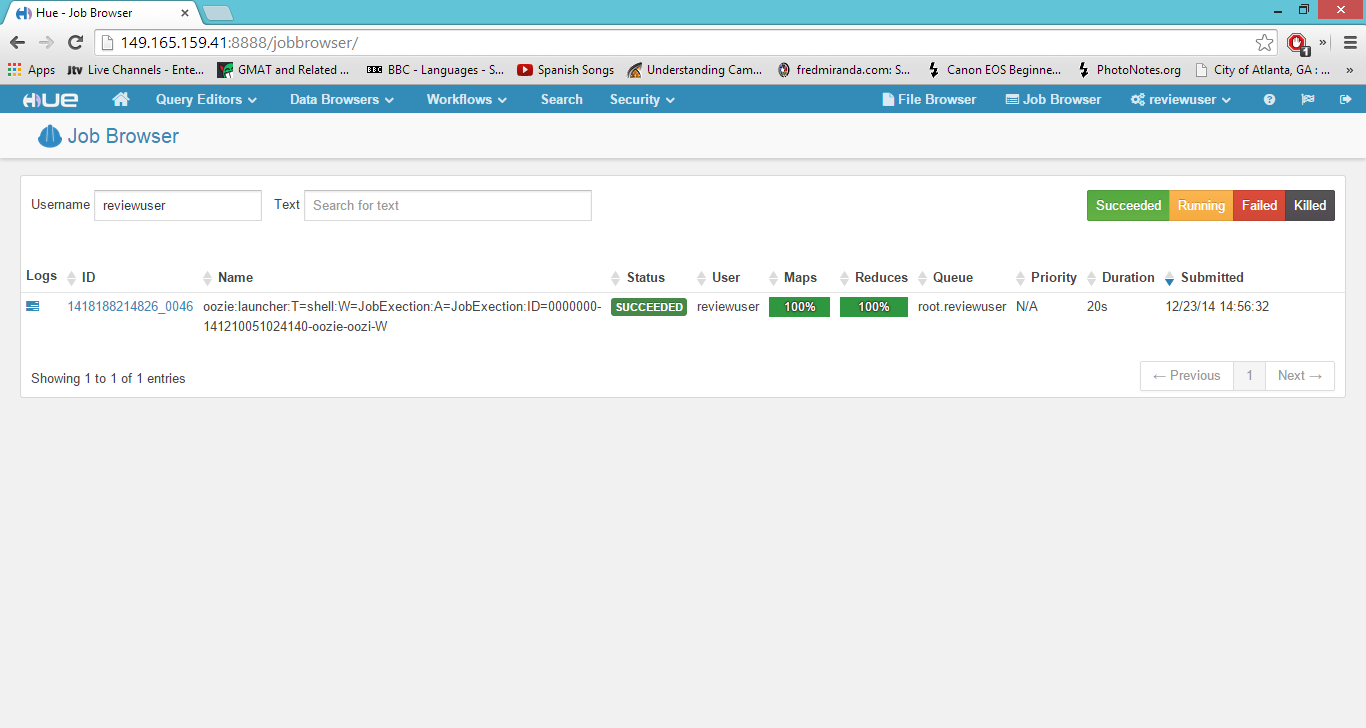
sudo -u hdfs hadoop jar ~/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar com.naveen.drivers.ItemVectorDriver 1 2

sudo -u hdfs hadoop jar ~/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar com.naveen.drivers.MergeMatrixItemDriver 1 2

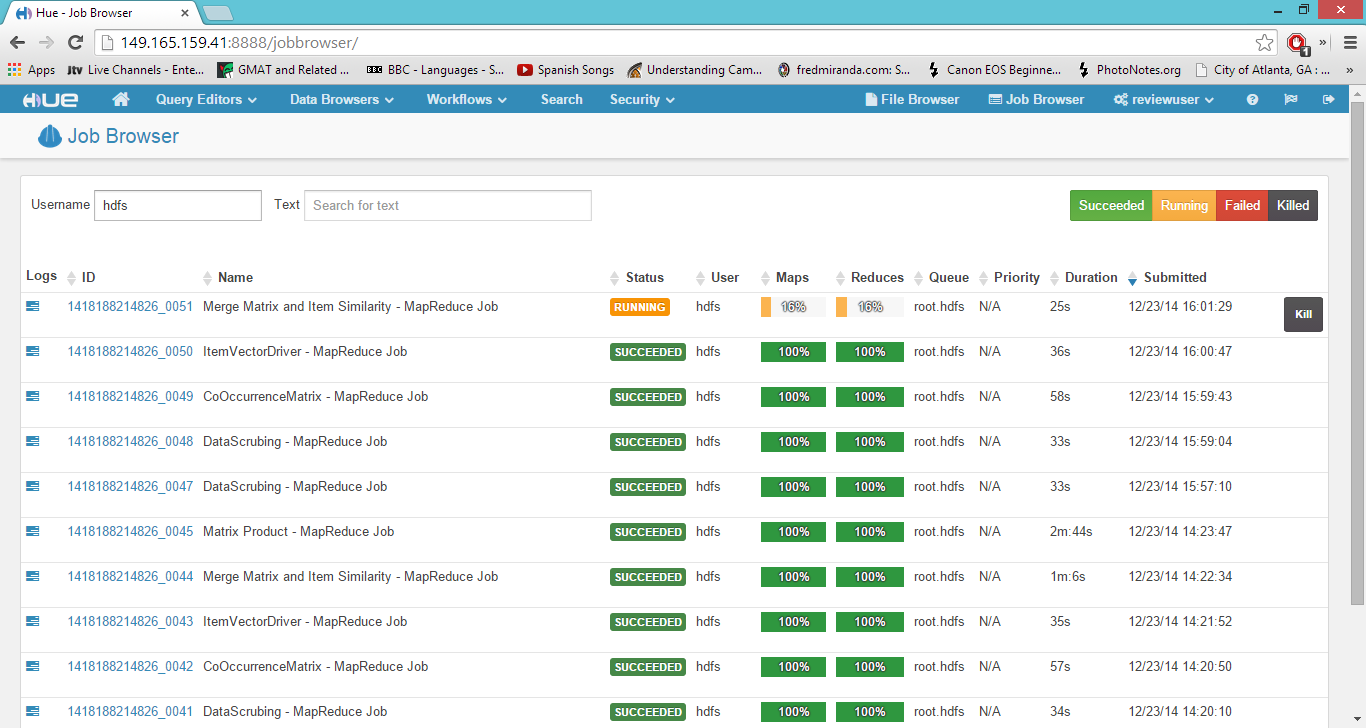
sudo -u hdfs hadoop jar ~/Jar/ItemBasedReco-0.0.1-SNAPSHOT.jar com.naveen.drivers.MatrixProductDriver 1 2

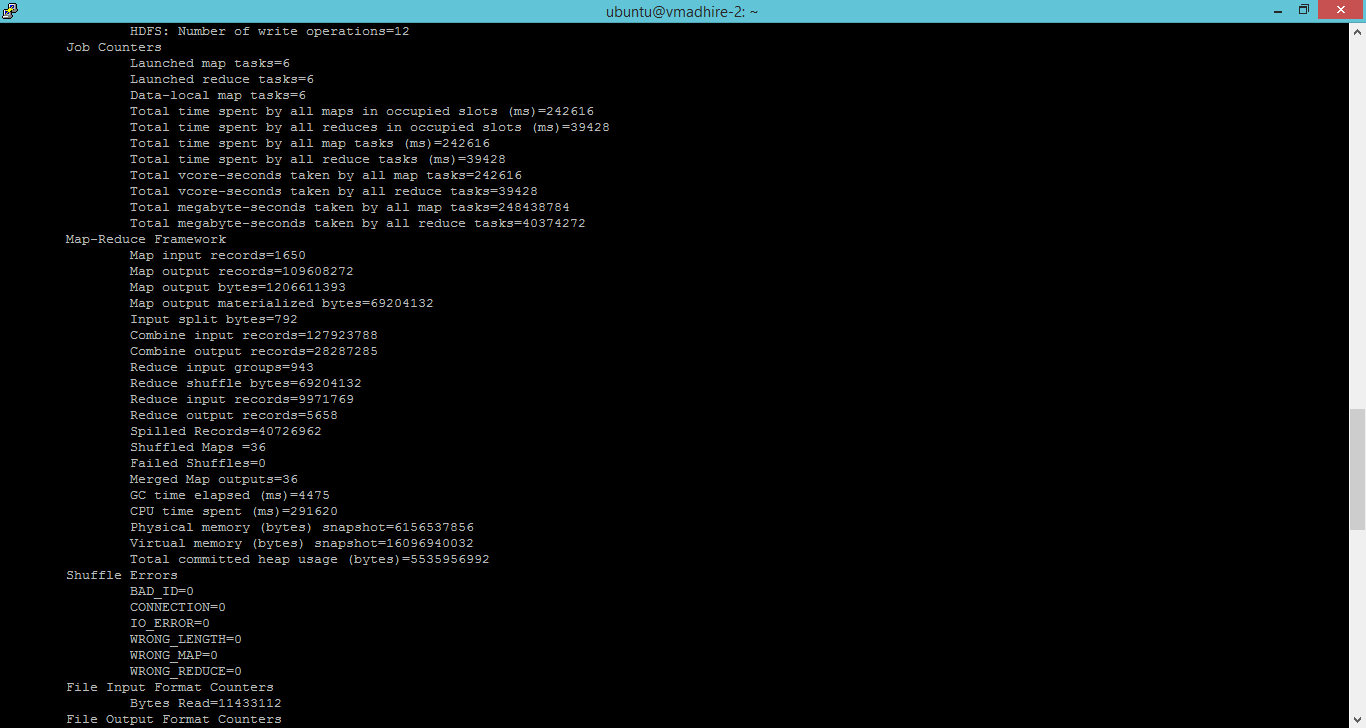
1. 5 jobs should be submitted by the above shellscript

The status of all the submitted jobs can be seen at the Job Browser tab. This Job Browser tab is for checking the status of Oozie submitted jobs.



To check the jobs submitted by the shell script, filter by “hdfs” in the username tag.



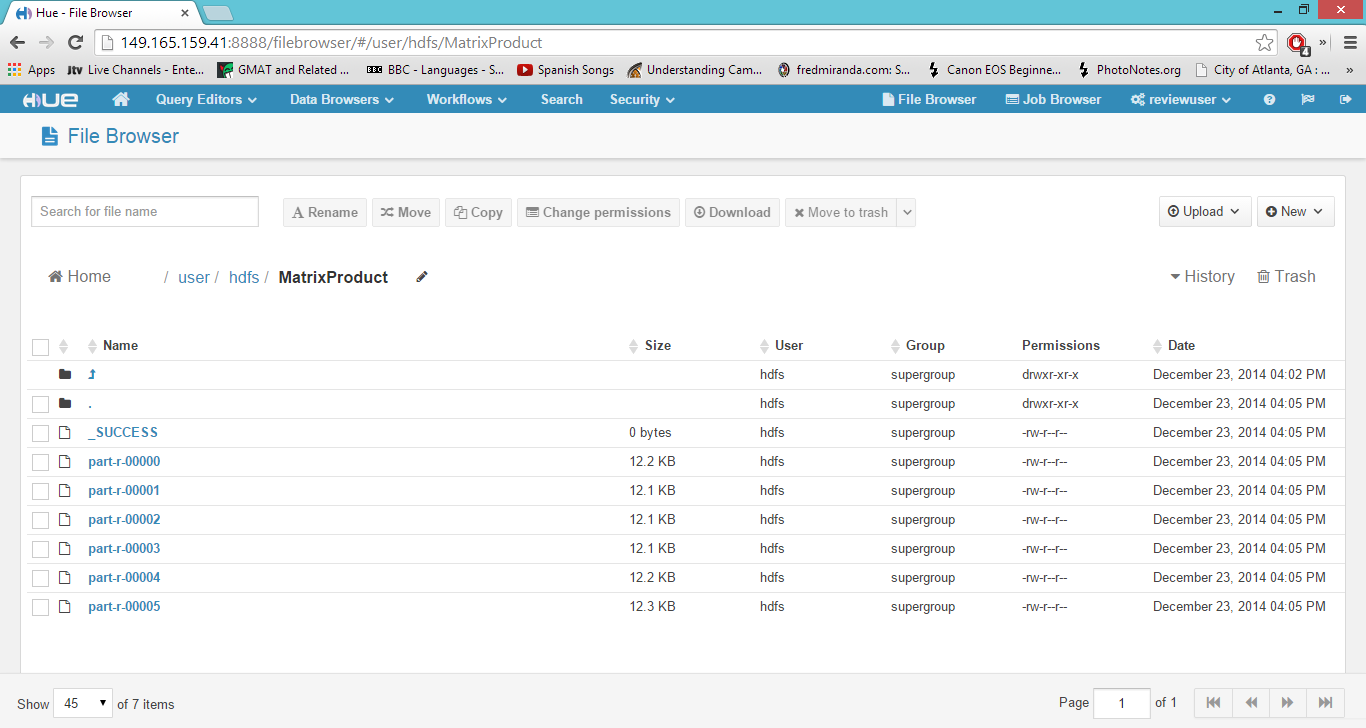


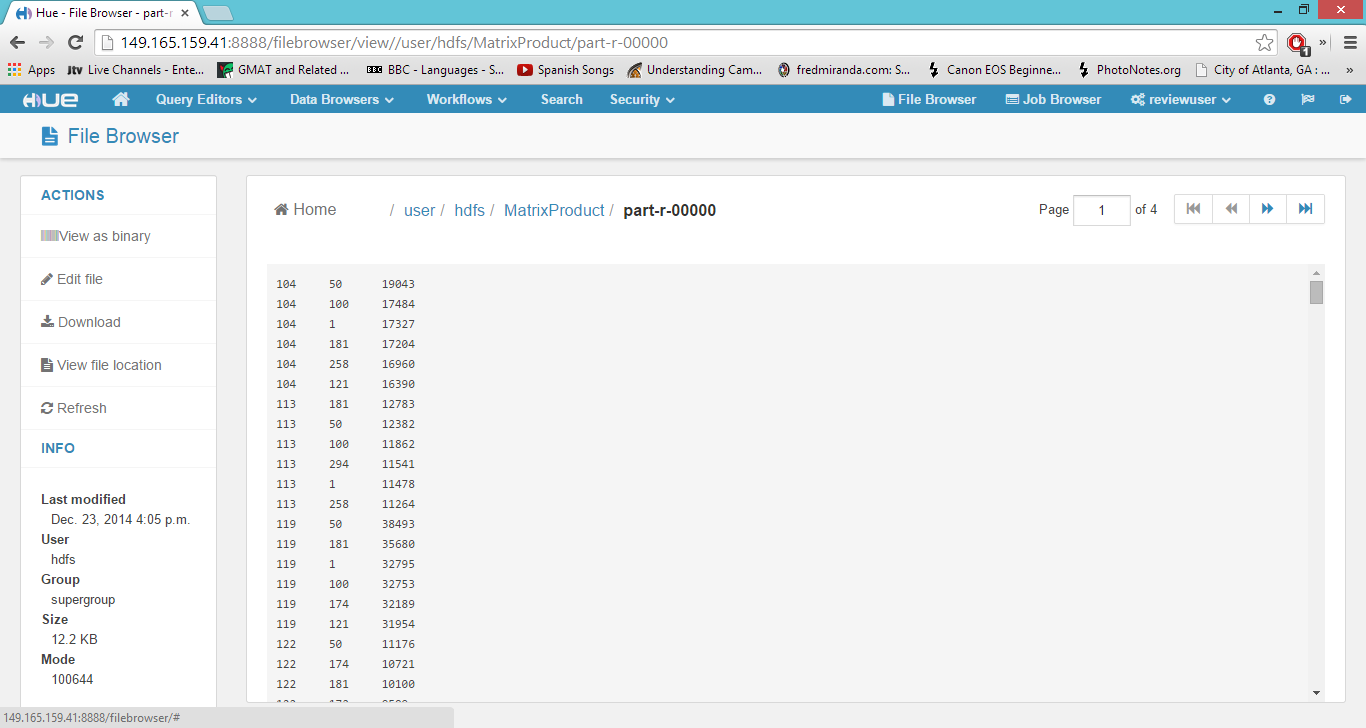
1. The final output file present in the hdfs location “/user/hdfs/MatrixProduct” location.

The output is a tab separated file. Consisting of top 5 movies recommended for each user based on their past ratings of the movies.

Details in output file as shown in the screen shot.

UserId, MovieId, Relative rating





1. Future work

The output file generated can be consumed downstream by a web application, reporting tool for using the recommendations.