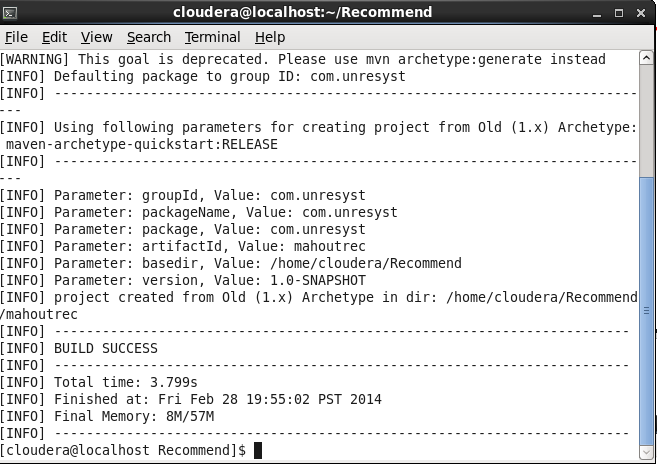
LAB 4

Name:Vishnu Praveen Gude

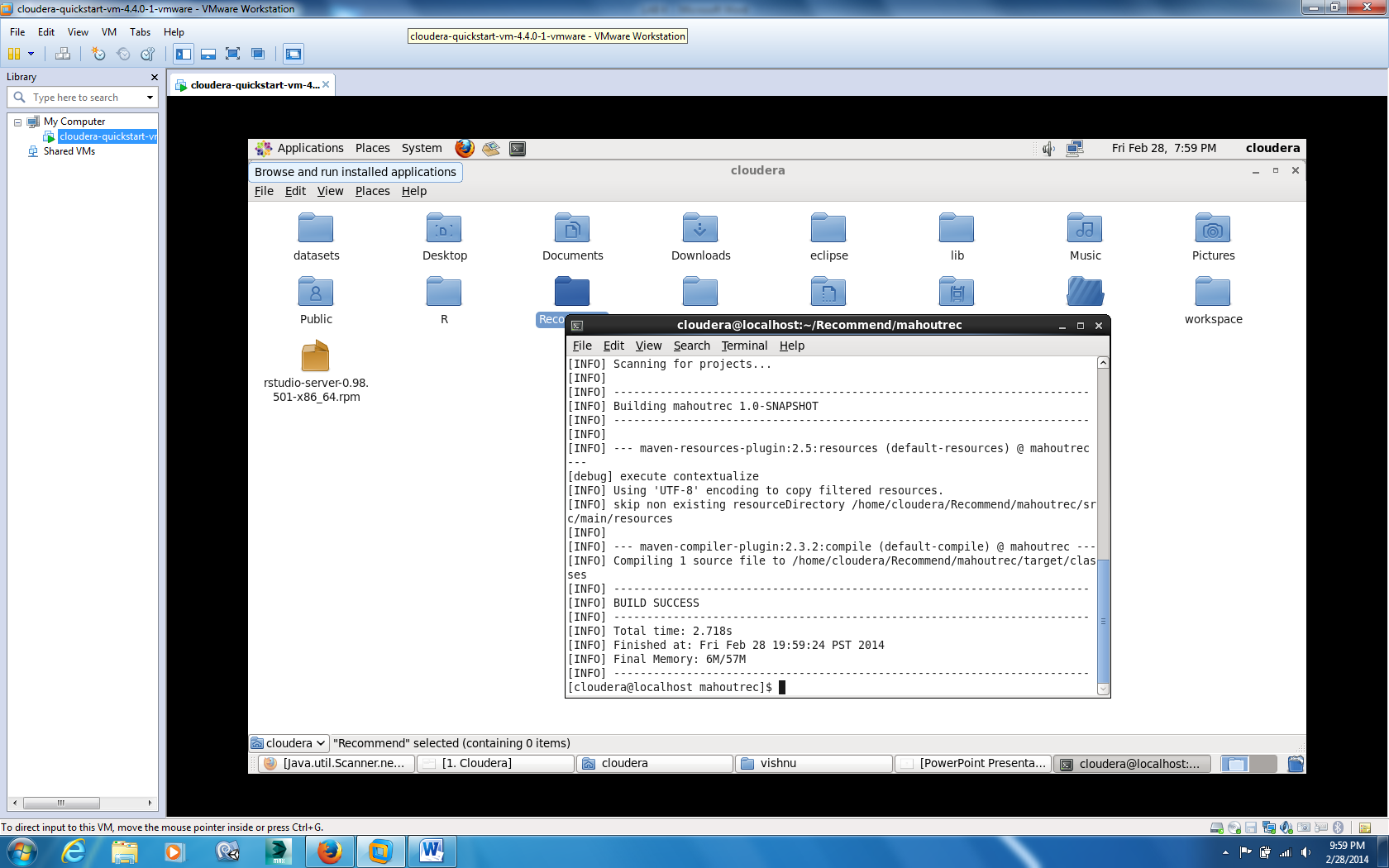
Id:16103389

Mahout/solr Recommendation System

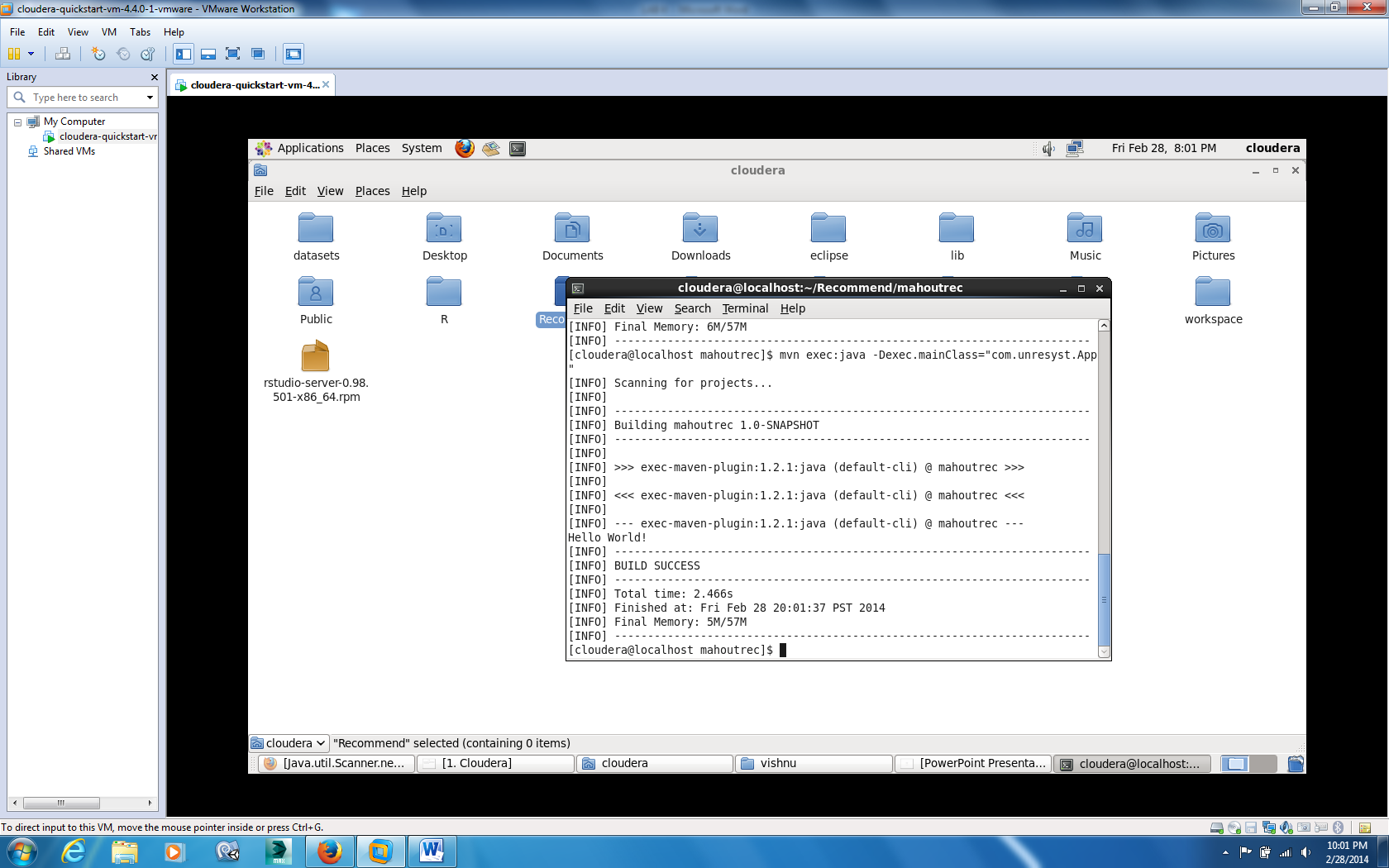
1.Create Empty mahout project name mahoutrec with package “com.unresyst”



Check with mvn compile



Hello world test



Modified pom.xml file and created datasets folder.

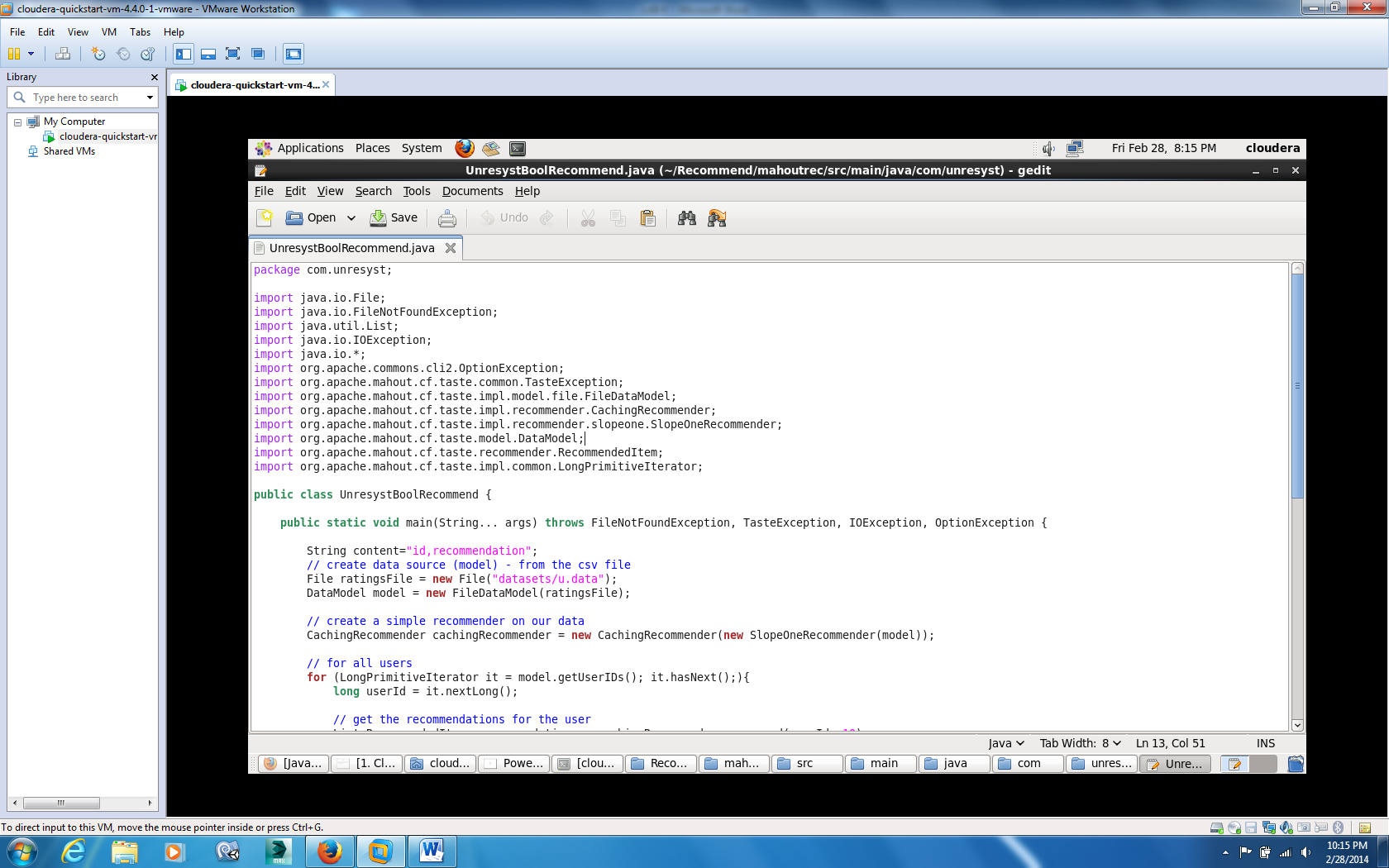
From grouplens.org I download 100K dataset

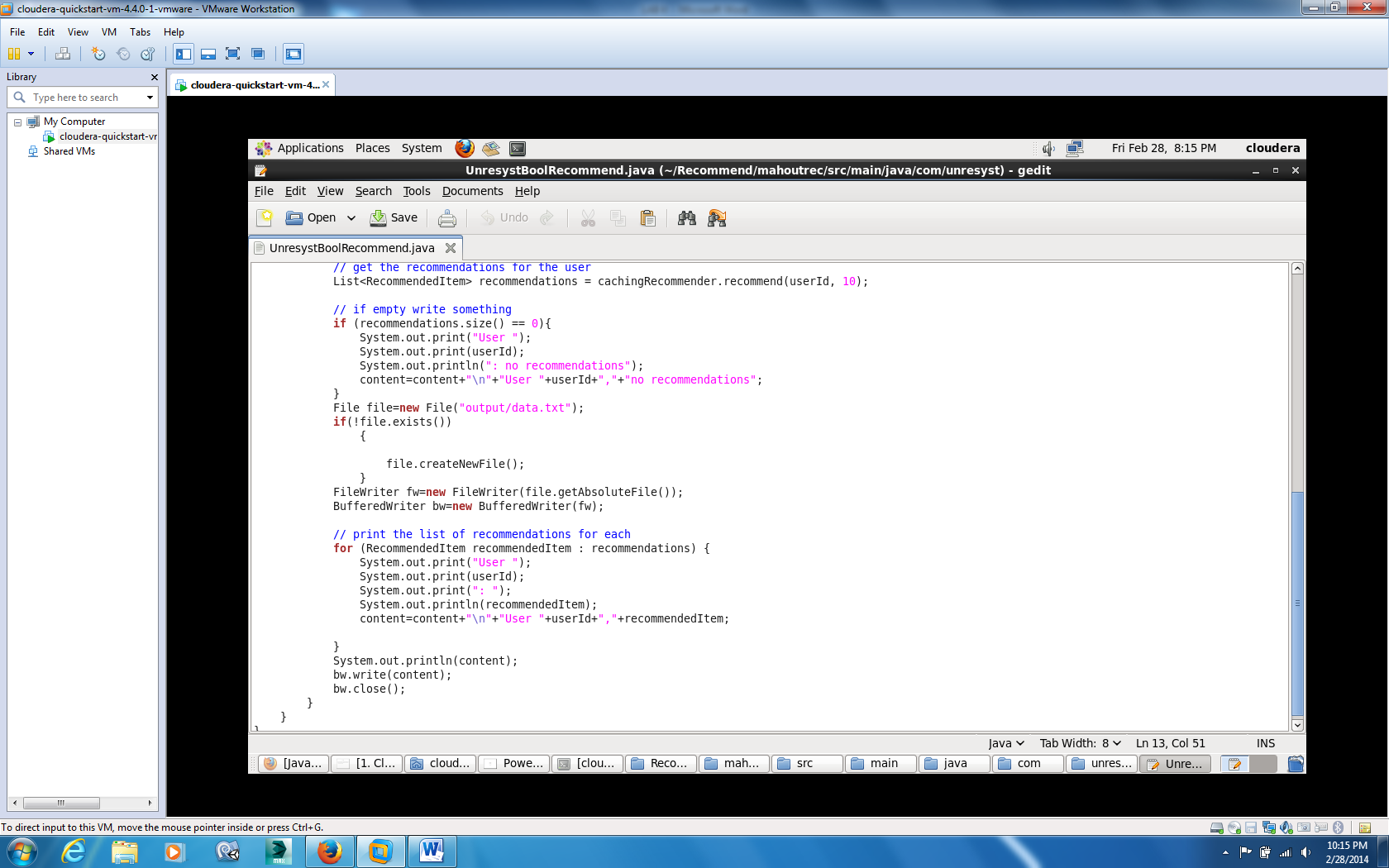
<http://grouplens.org/datasets/movielens/>

I placed u.data into datasets folder and I created output folder under mahoutrec.

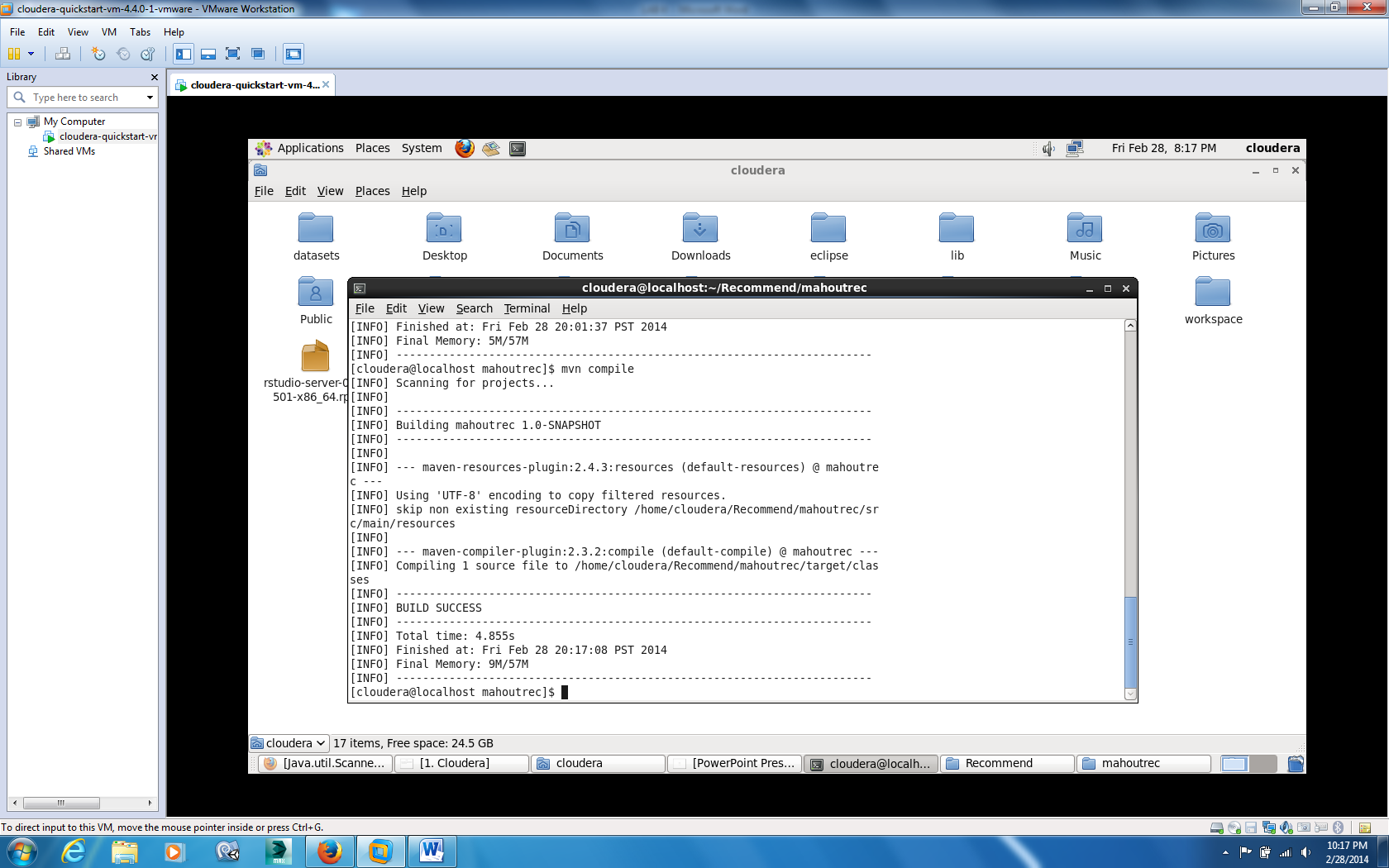
I modified UnresystBoolRecommend.java to generate a recommendation based on itemId and save the results to data.txt file in output folder….

Final UnresystBoolRecommend.java

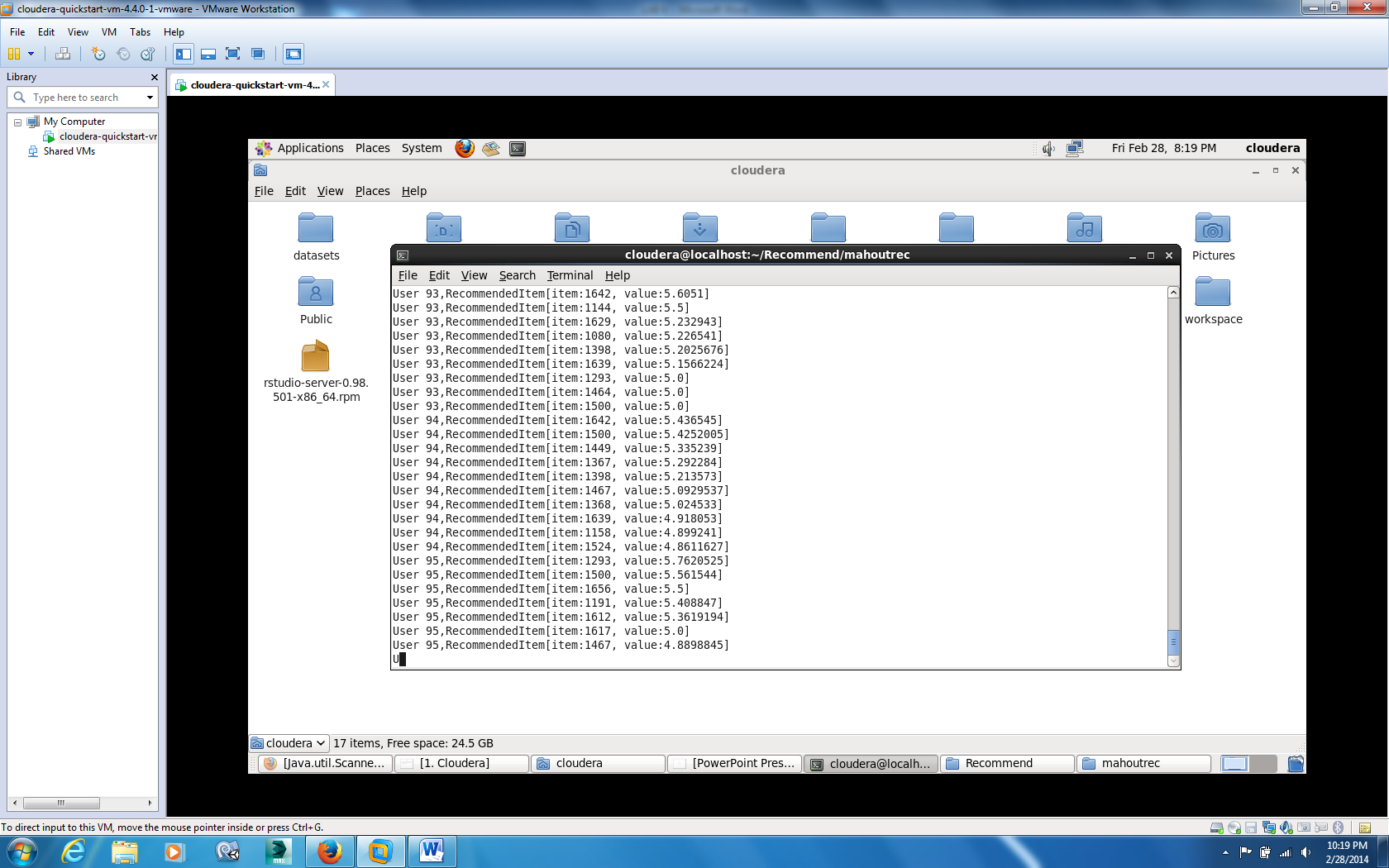




Now compile UnresystBoolRecommend.java file

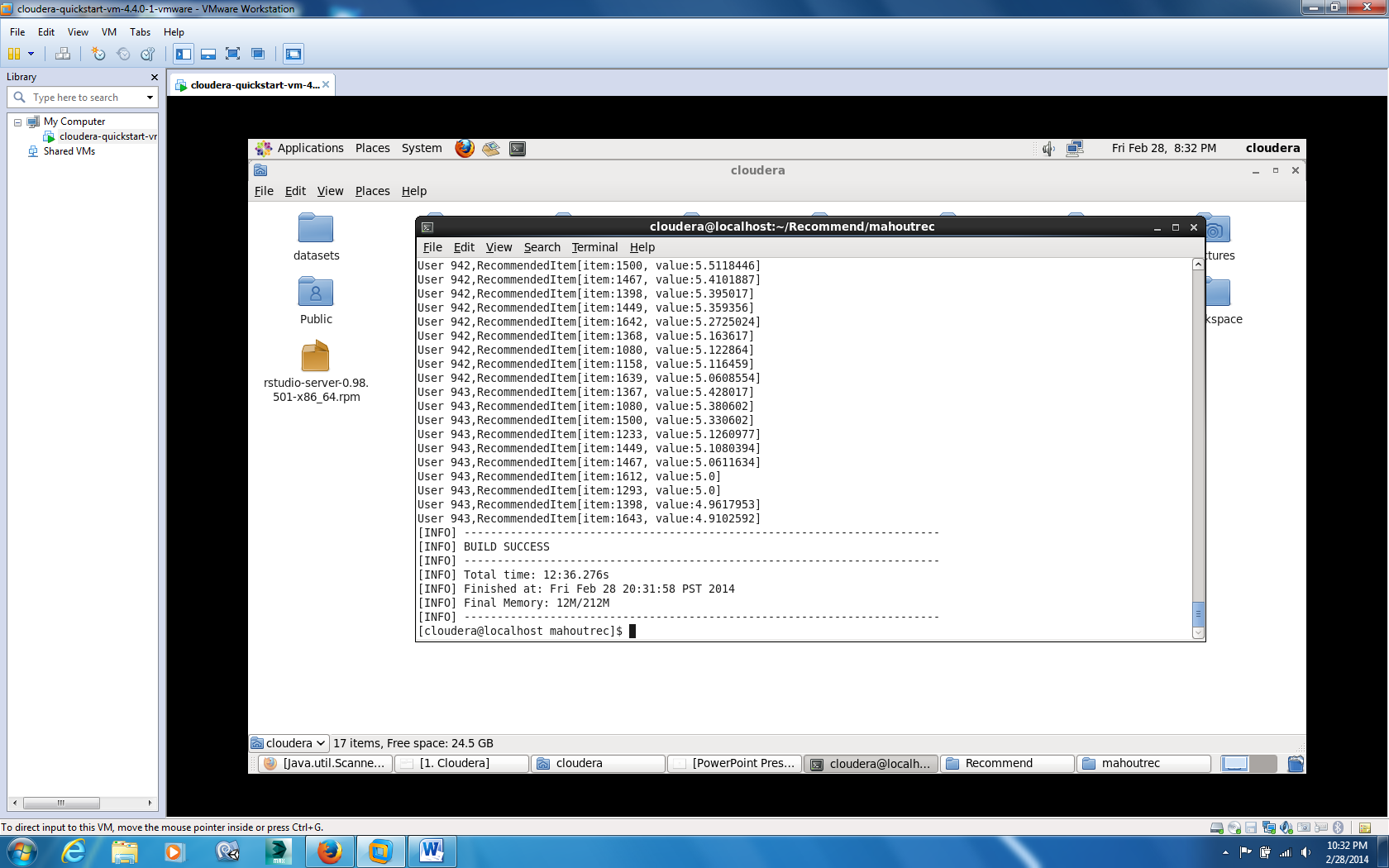


Now execute the main function

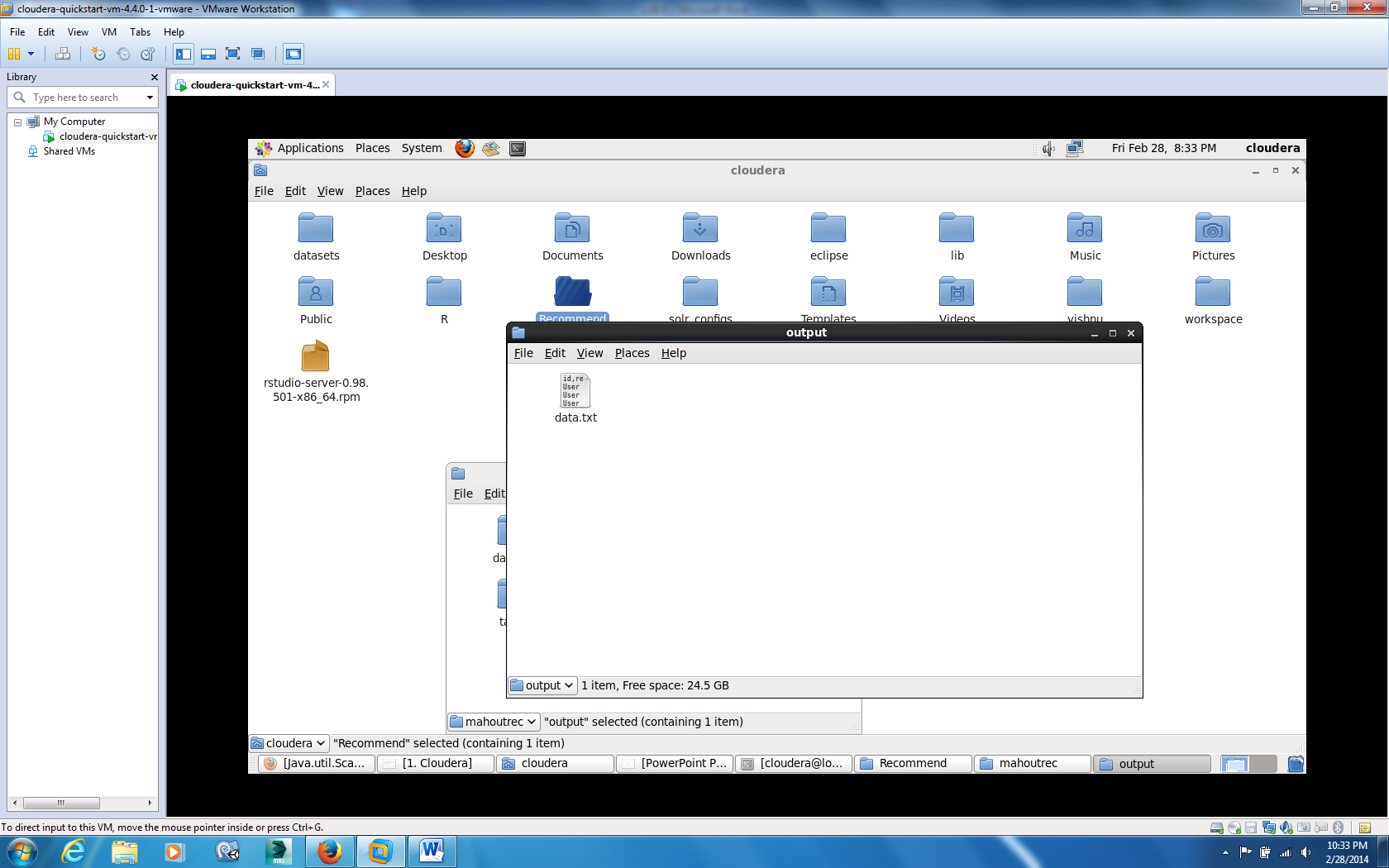


It will take some time because it’s a 100K dataset….

Now its succeeded



It should generate data.csv file in output folder….

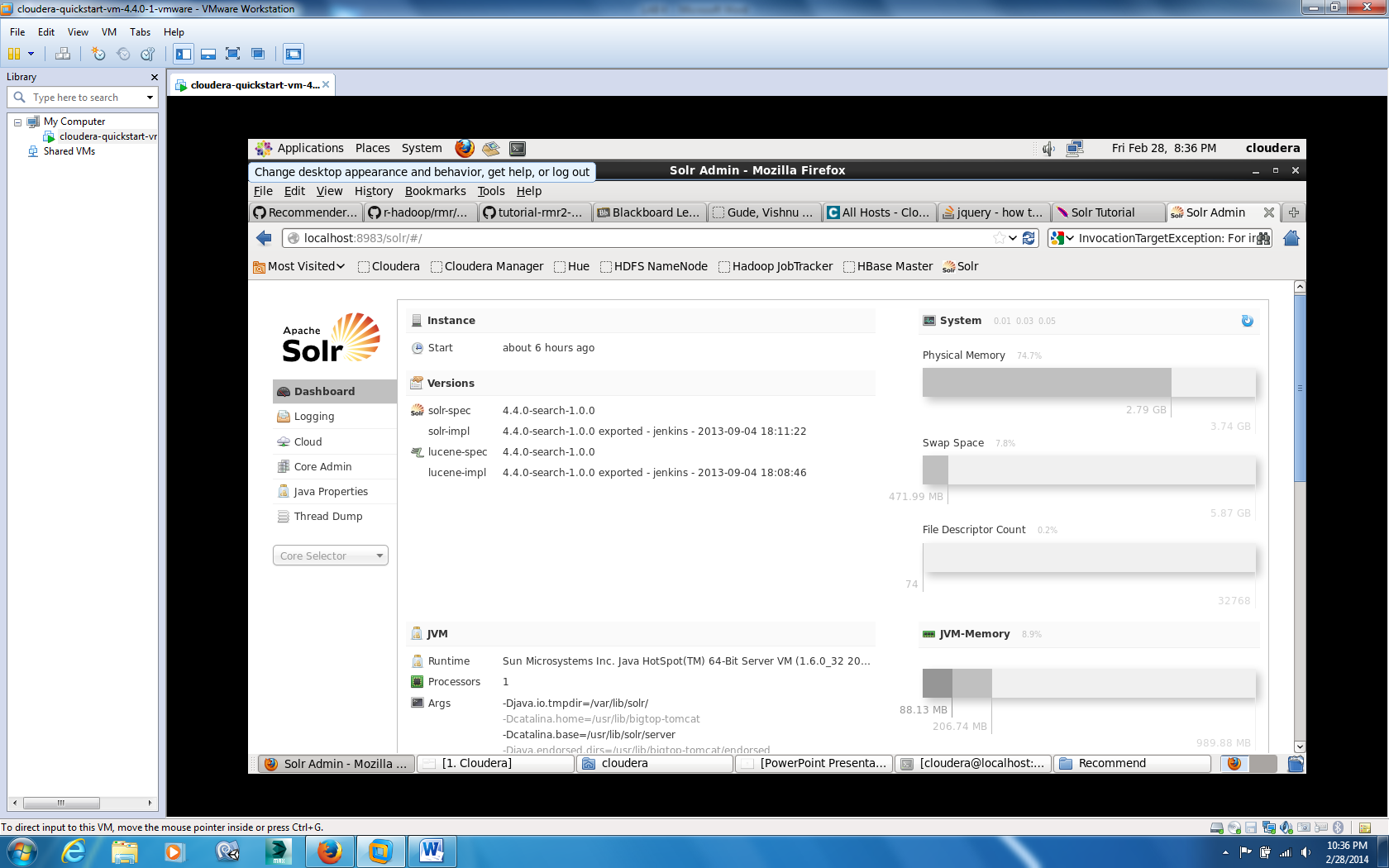


We need to push this mahout output to solr…..

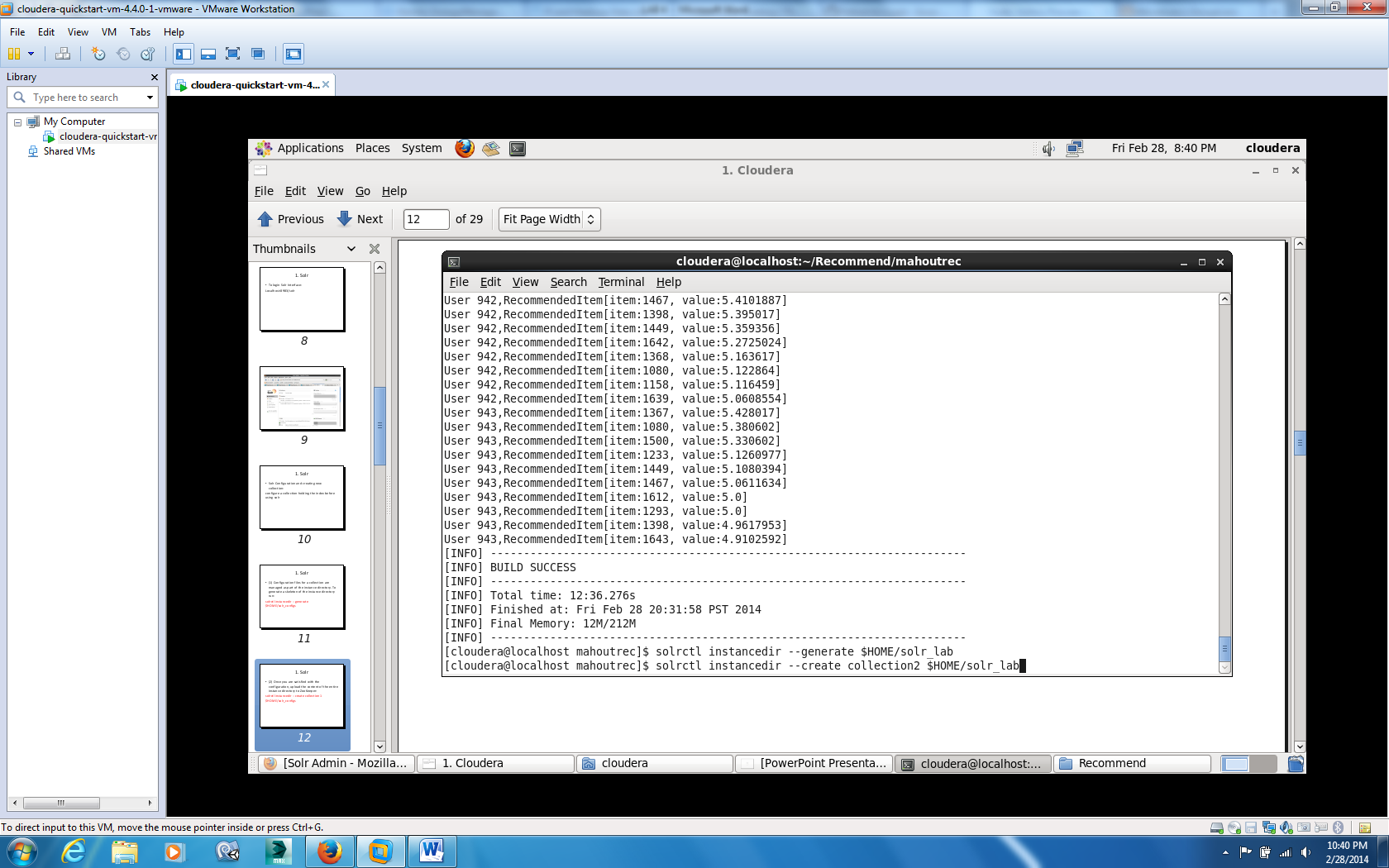
So first create solr project and collection…

Check solr panel first to make sure solr installation

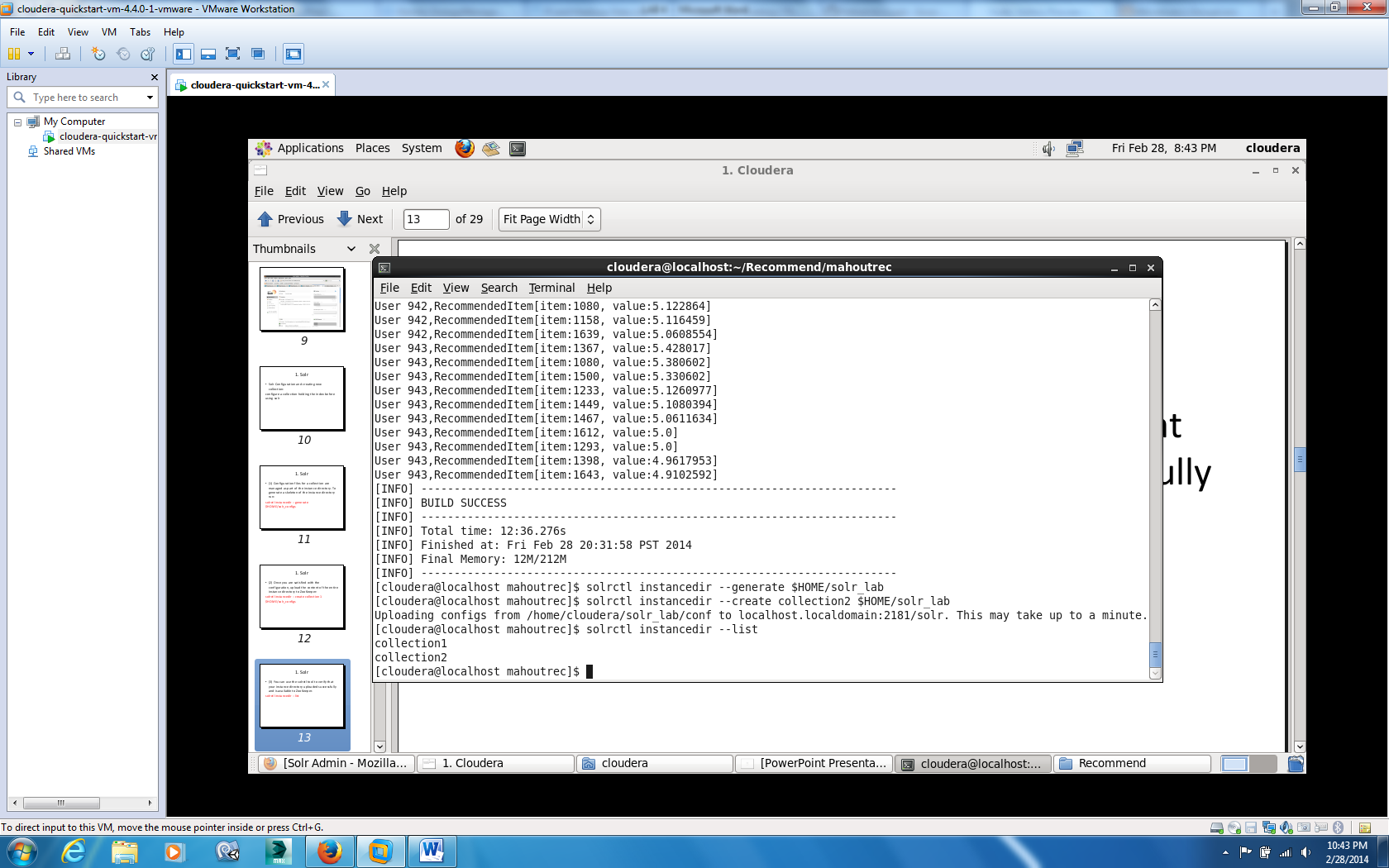
Localhost:8983/solr



Now,create the instance directory for configuration files for collection and move the content to zookeeper

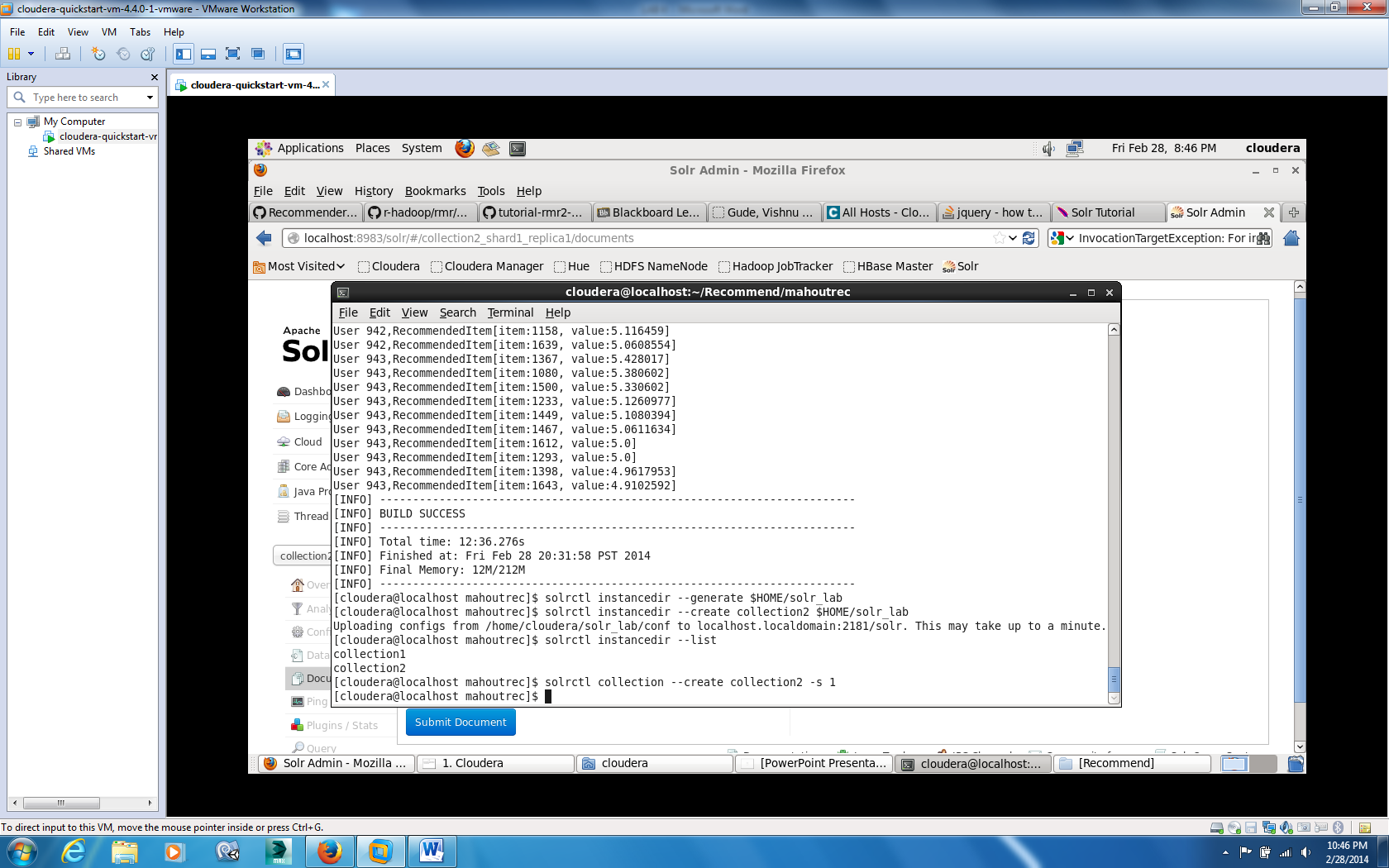


Now check instancedir is successfully uploaded to zookeeper



Yes we have collection2 in zookeeper

Now create solr collection with same name



Now check the solr admin panel this collection should be displayed.



Now its time to move mahout output to solr….But we need to do some steps before we do that…

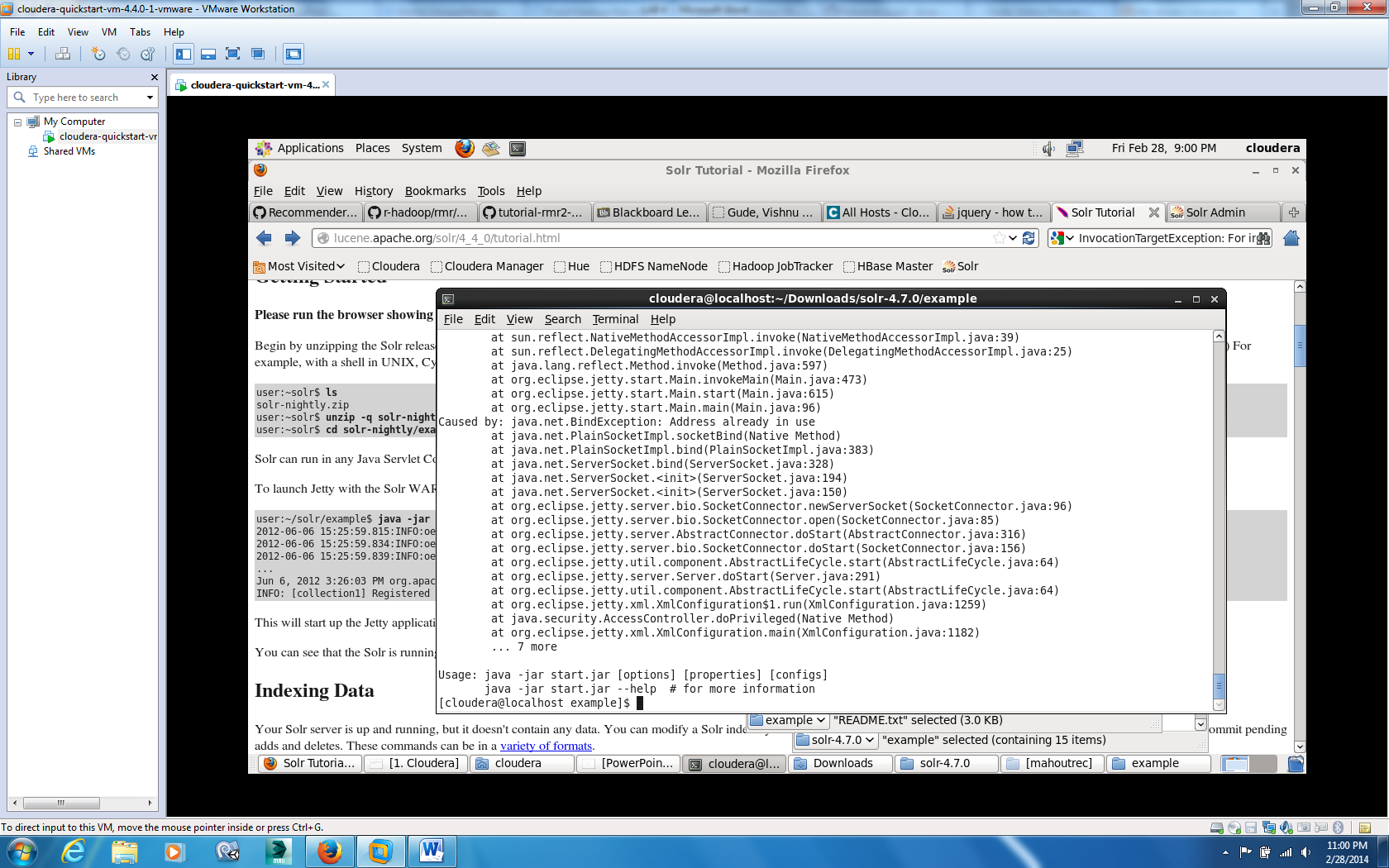
Download solr directory to your local machine from this url

<http://www.apache.org/dyn/closer.cgi/lucene/solr/4.7.0>

Now unzip the folder and move to solr location….you will find start.jar file under example folder

Run it by using command

Java –jar start.jar



I got error showing already address in use because I already started solr server before…..

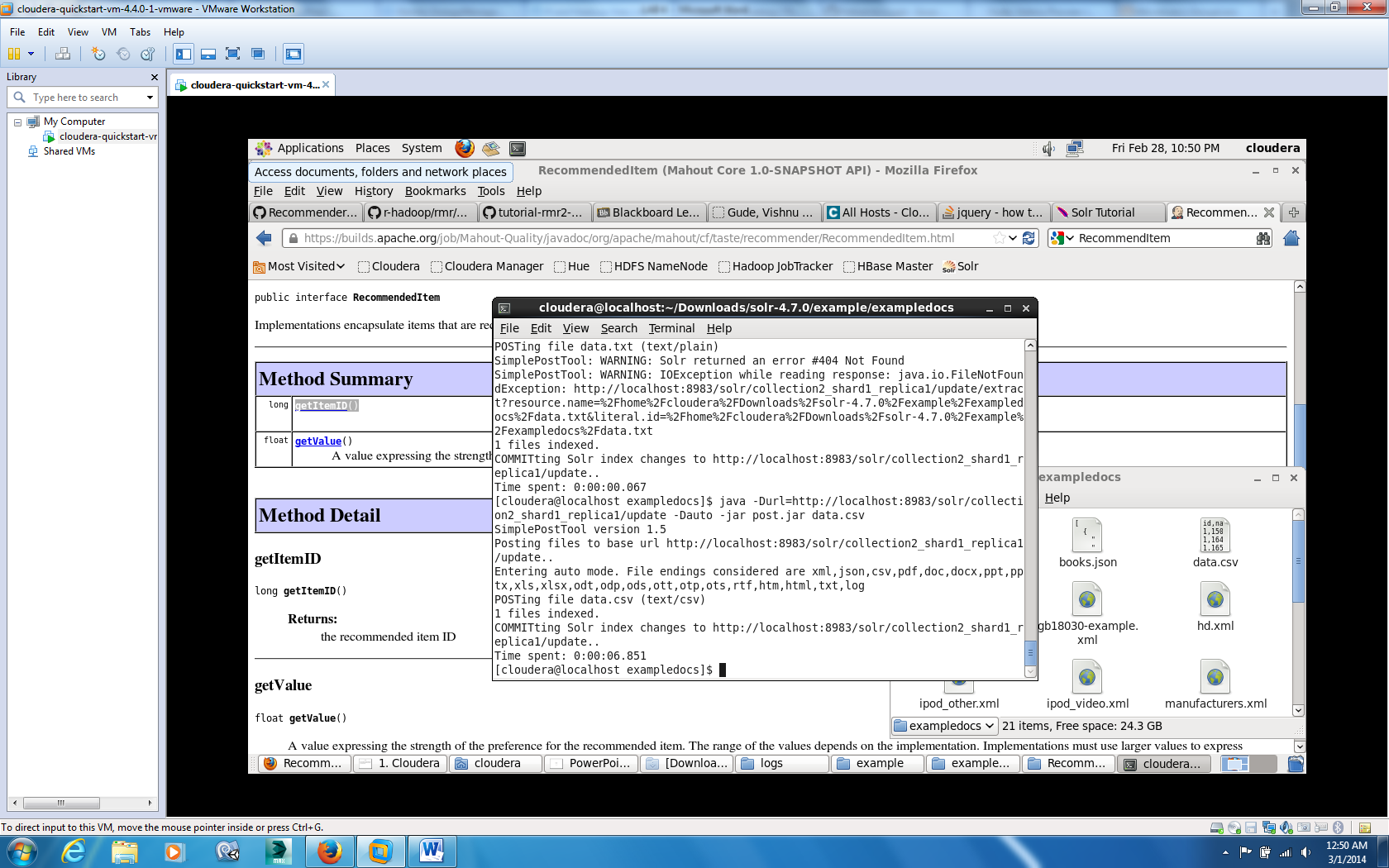
Now we need to push mahout output data.txt file to solr collection2

Copy the data.txt file under mahout output folder to solr-4.7.0/example/exampledocs/ folder

Now run the post.jar file under exampledocs in solr-4.7.0/example/exampledocs/ folder

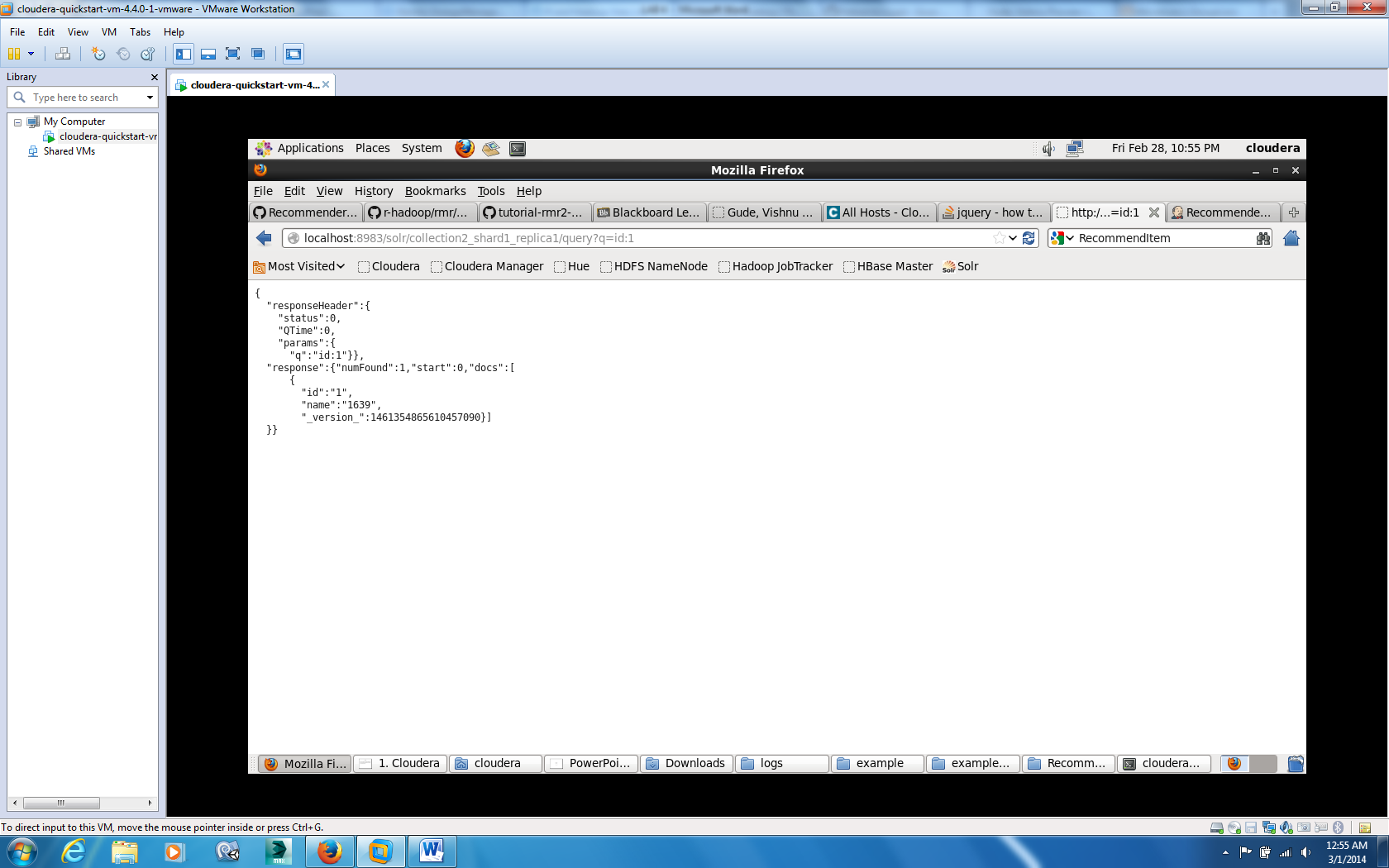
Now push file/files using command

Java –Durl=http://localhost:8983/solr/collection2\_shard1\_replica1/update -Dauto -jar post.jar data.csv



Now we can check Rest service by checking url

<http://localhost:8983/solr/collection2_shard1_replica1/query?q=id:1>



Similarly

<http://localhost:8983/solr/collection2_shard1_replica1/query?q=id:123>

