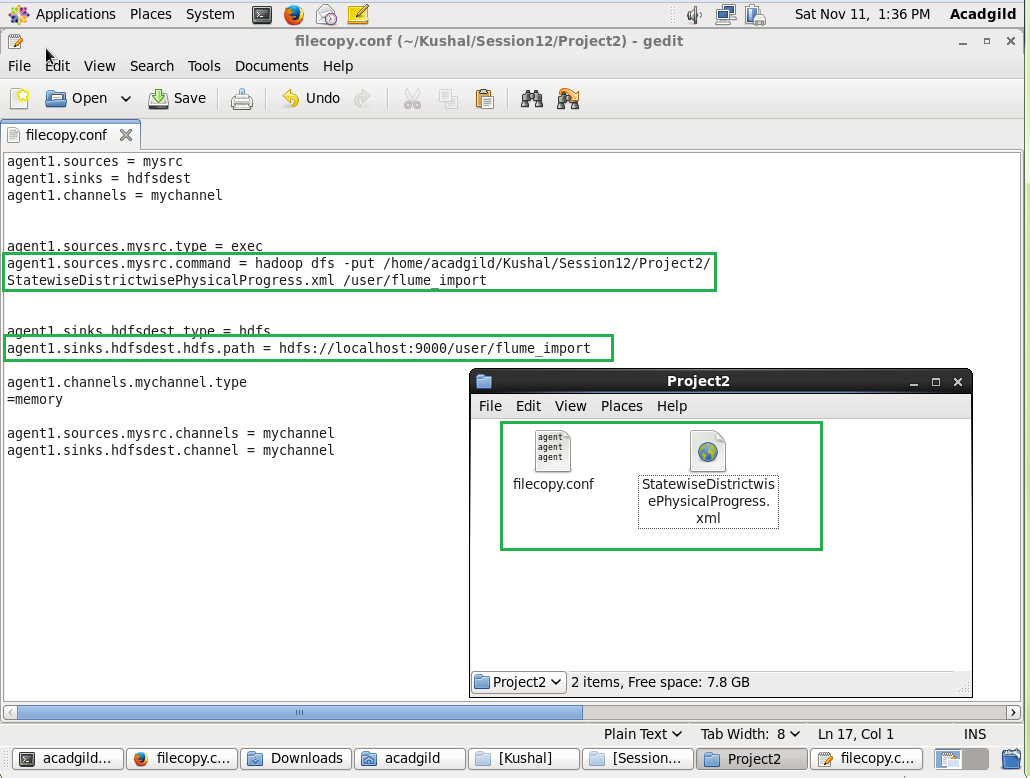
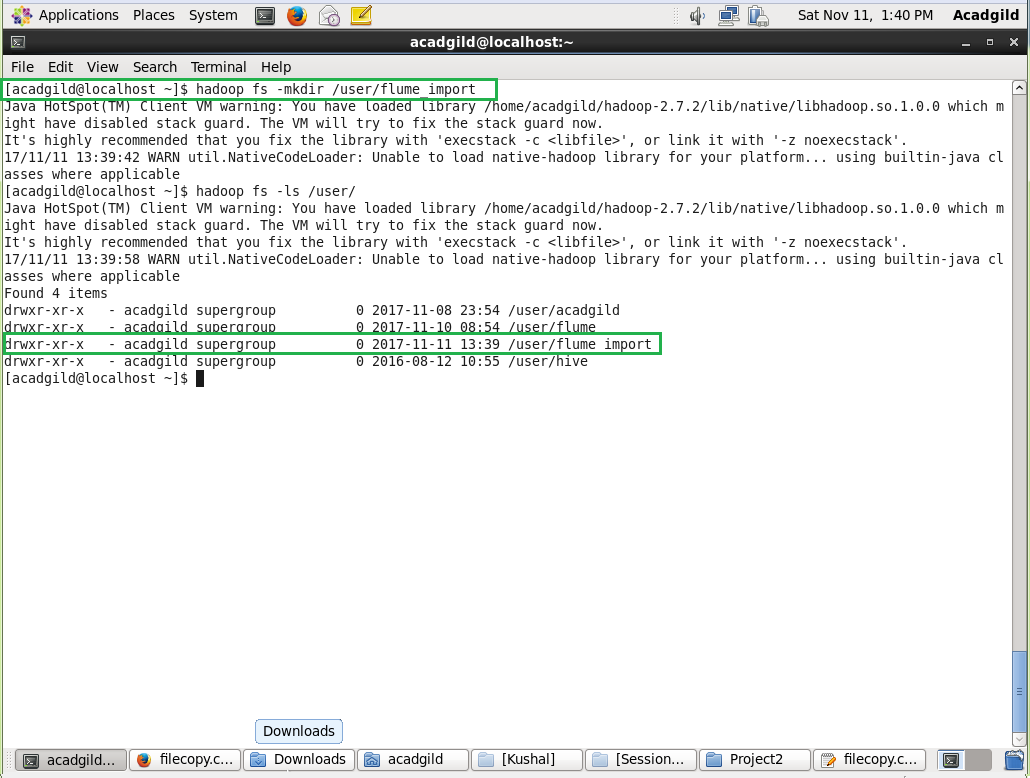
**Session 12 Project 2**

**Step 1:** Copy dataset from local file system to HDFS using flume.

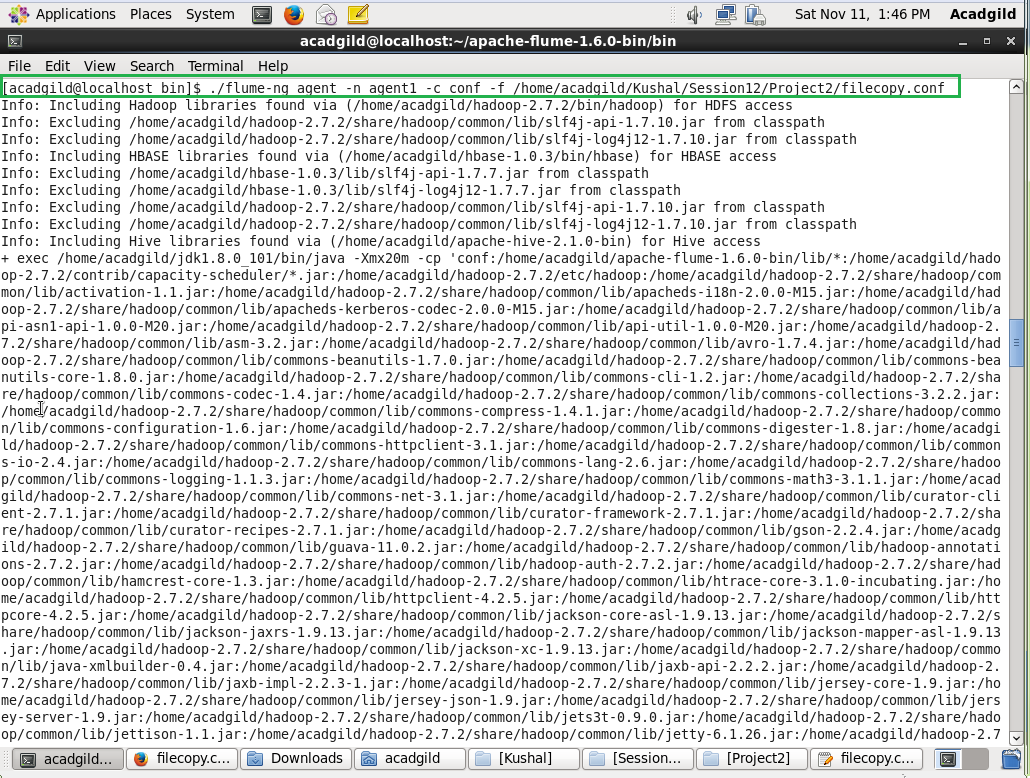
* filecopy.conf file is updated as per the local paths



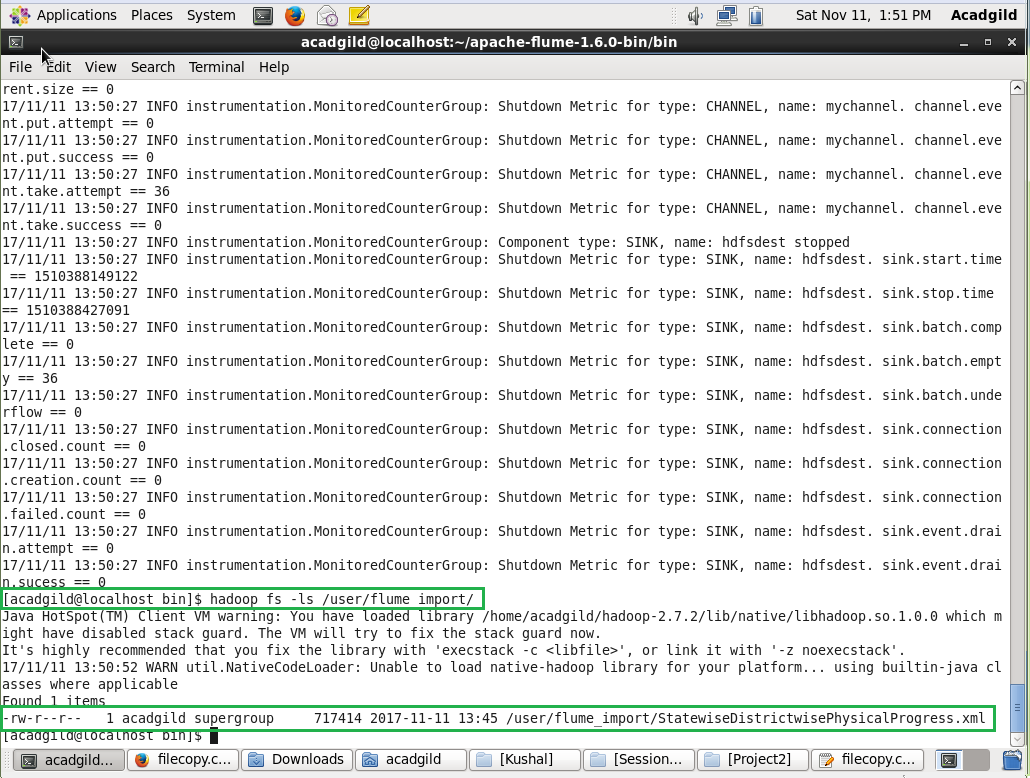
* Now Add the folder which is mentioned in the conf file to the hdfs



* Run the flume command to import the xml file to hdfs by giving the conf file path in command



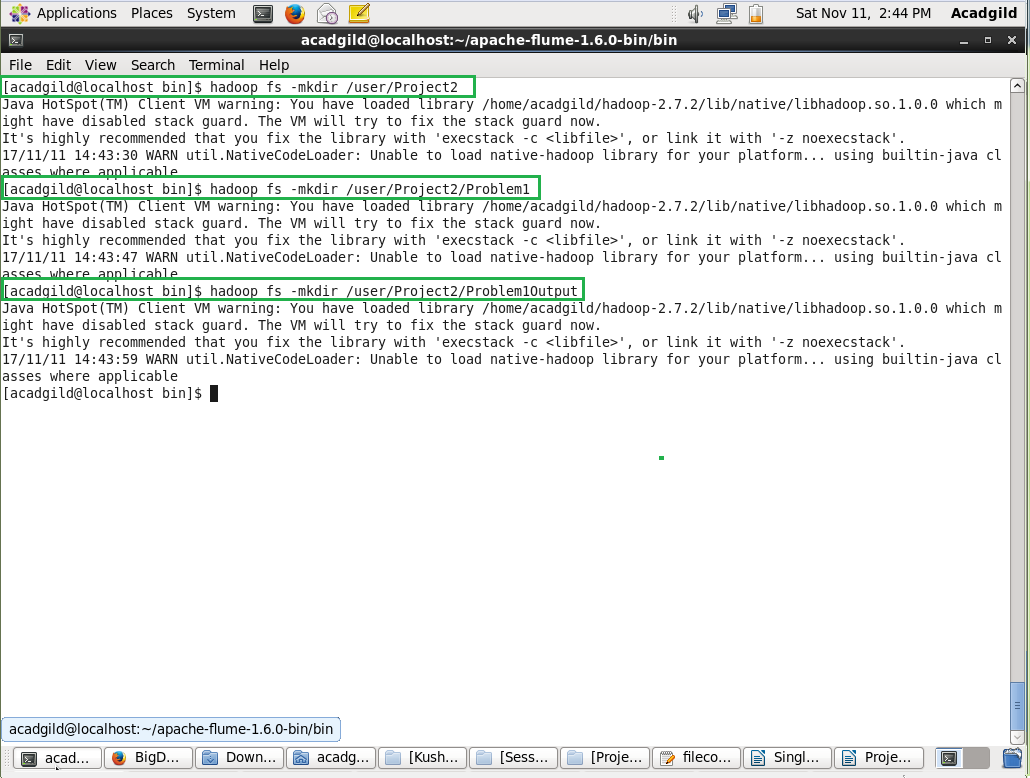
* File has been imported



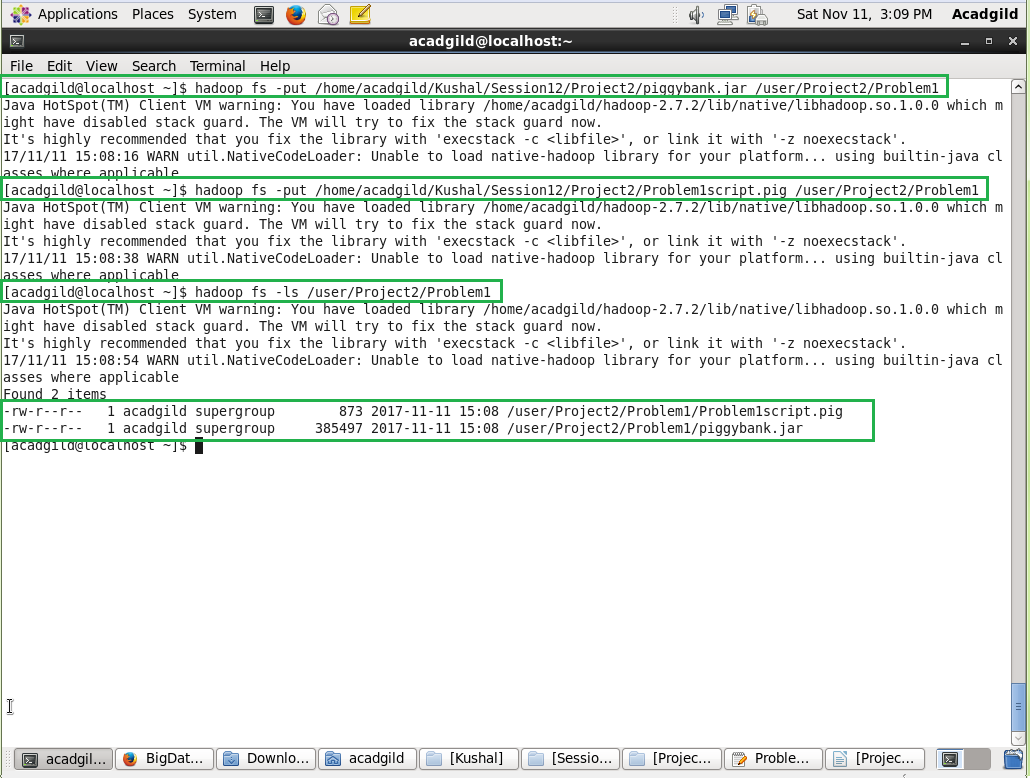
**Step 2:** Input file is in the XML format use Map reduce or pig to parse the data and get the results for the below problem statements.

**Problem 1:** Find out the districts who achieved 100 percent objective in BPL cards. Export the results to mysql using sqoop

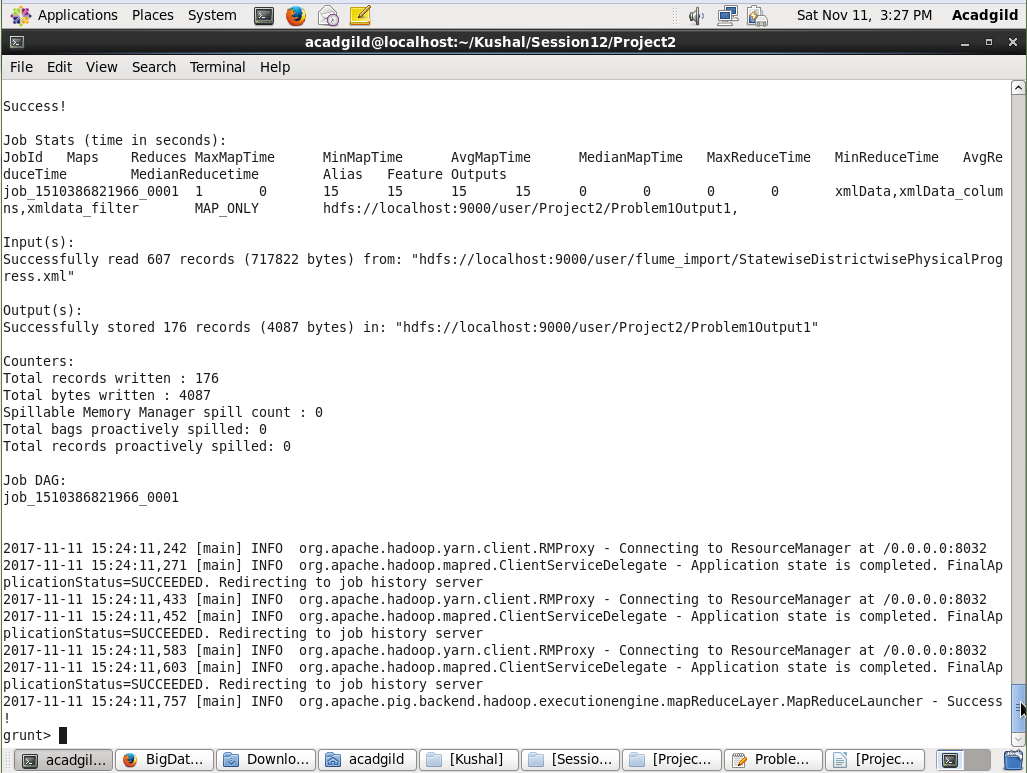
* Create the folders as shown in the below screenshot in HDFS



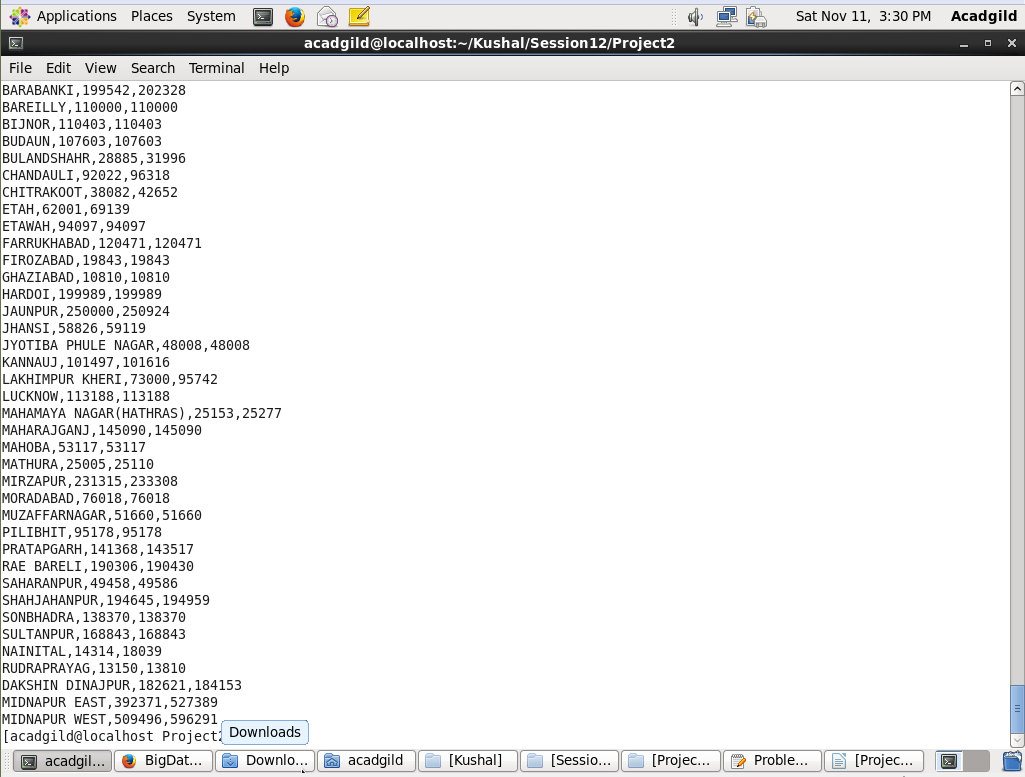
* Add the pig script and the piggybank.jsr to the /user/Project2/Problem1 to execute



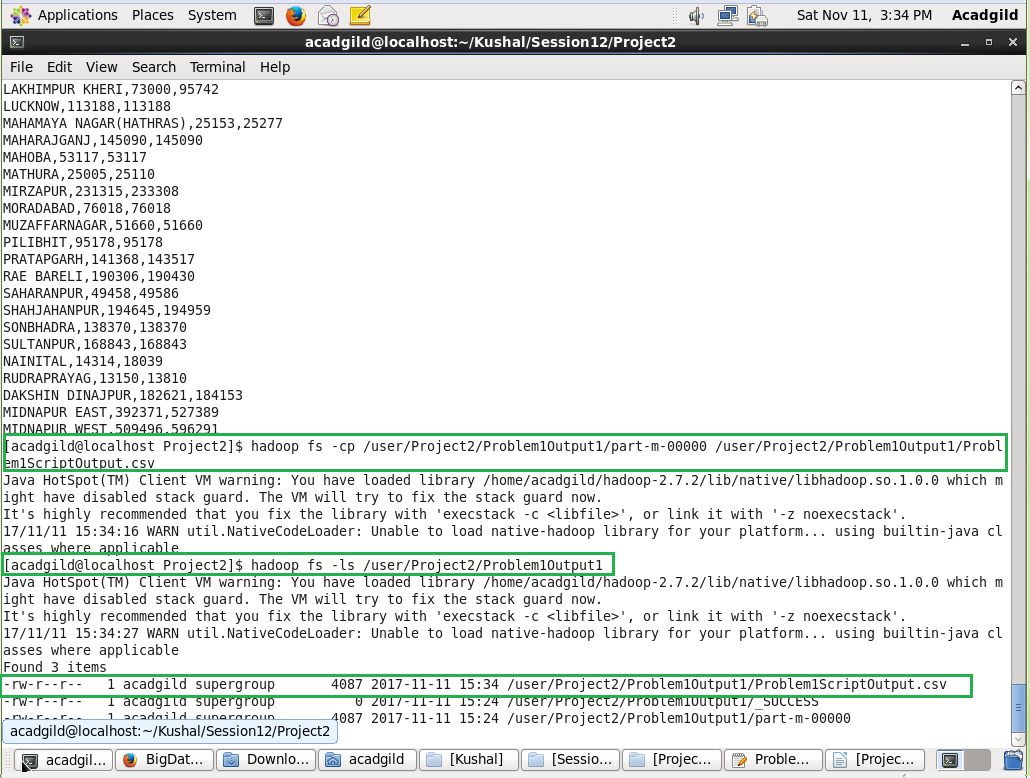
* Logon to grunt shell to execute the pig script and execute



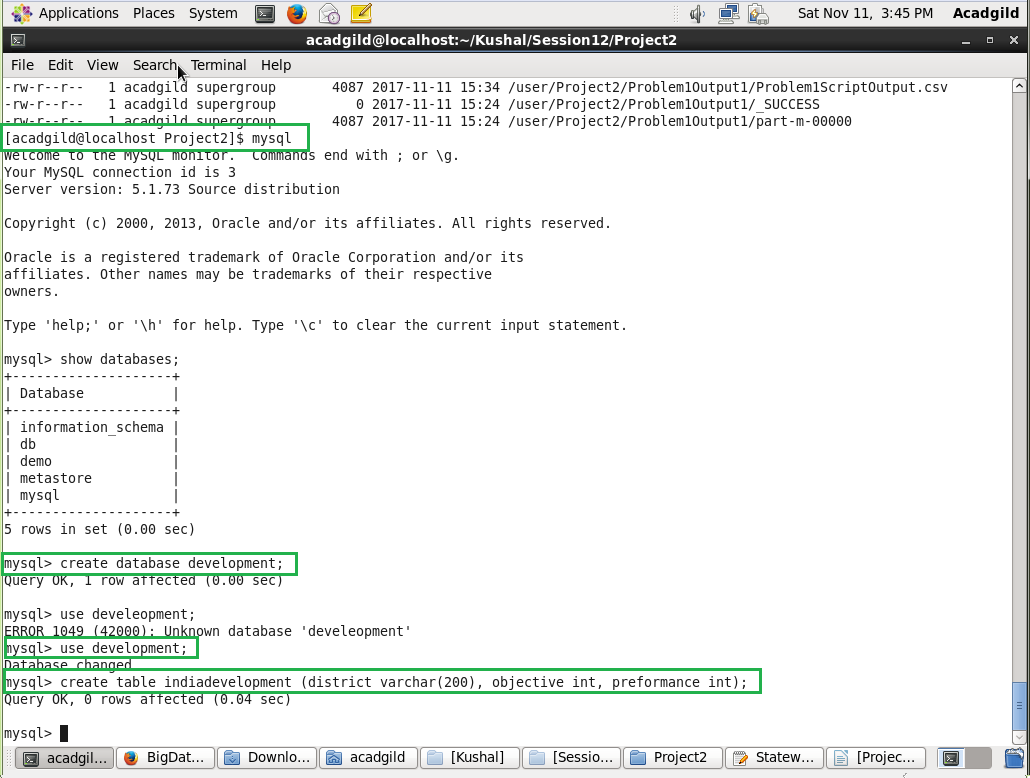
* Output from HDFS



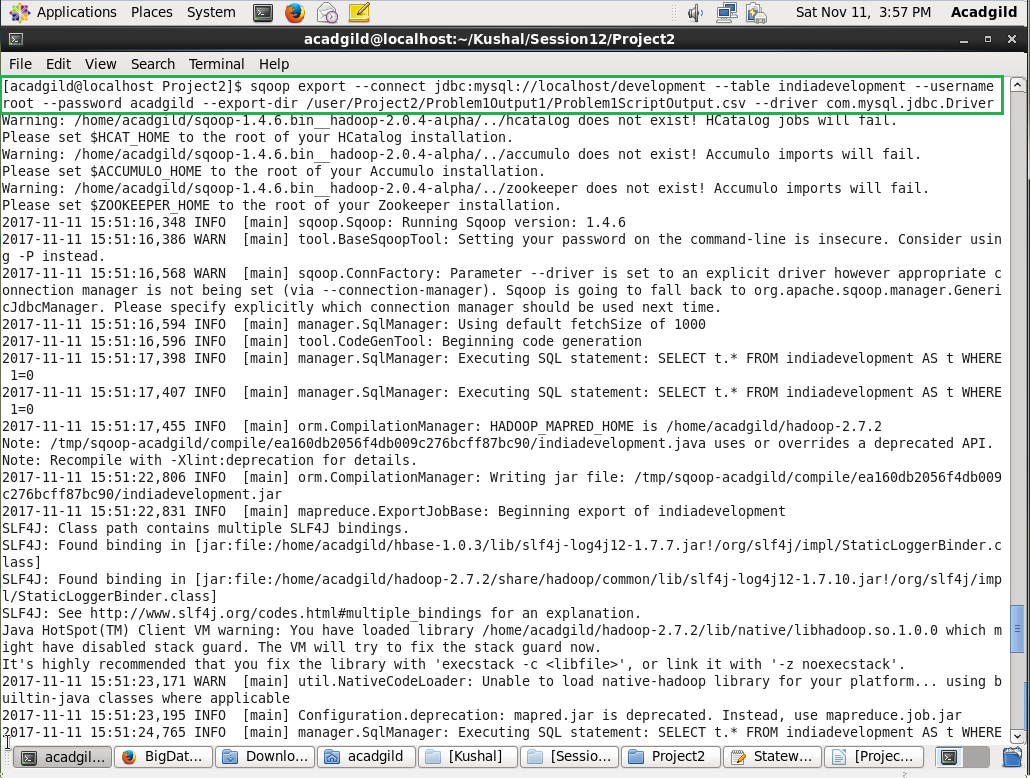
* Convert the output file to .csv to import to MySQL

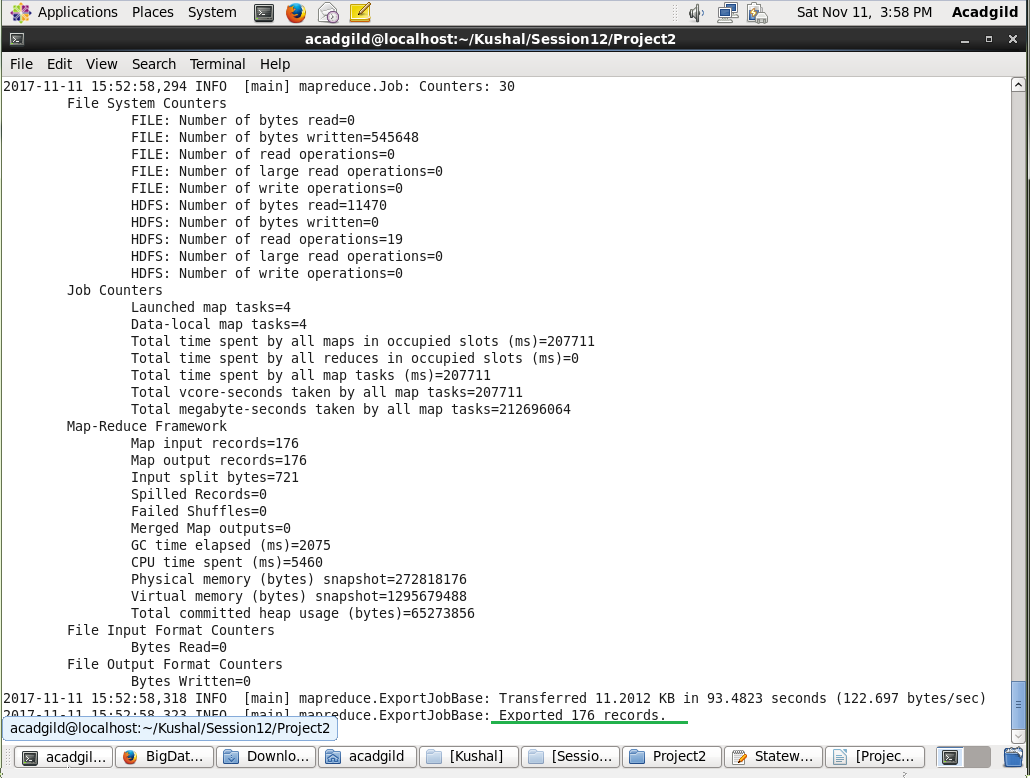


* Logon to MySQL and create a table for storing the .csv data

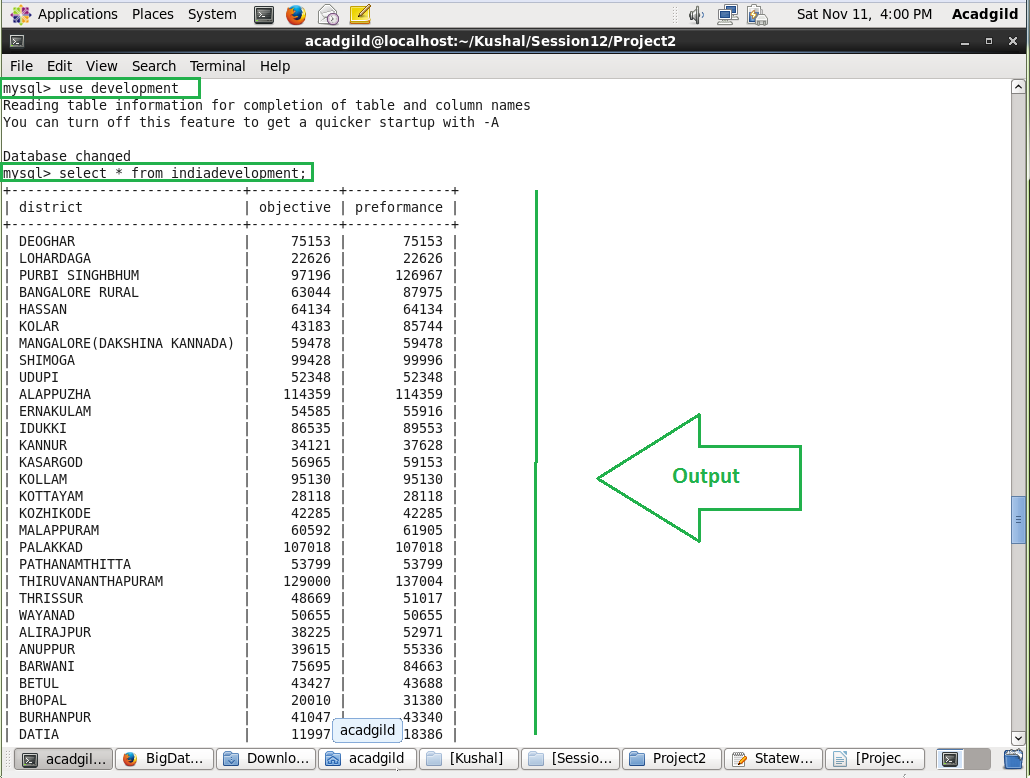


* Running sqoop command to export .csv data to MySQL



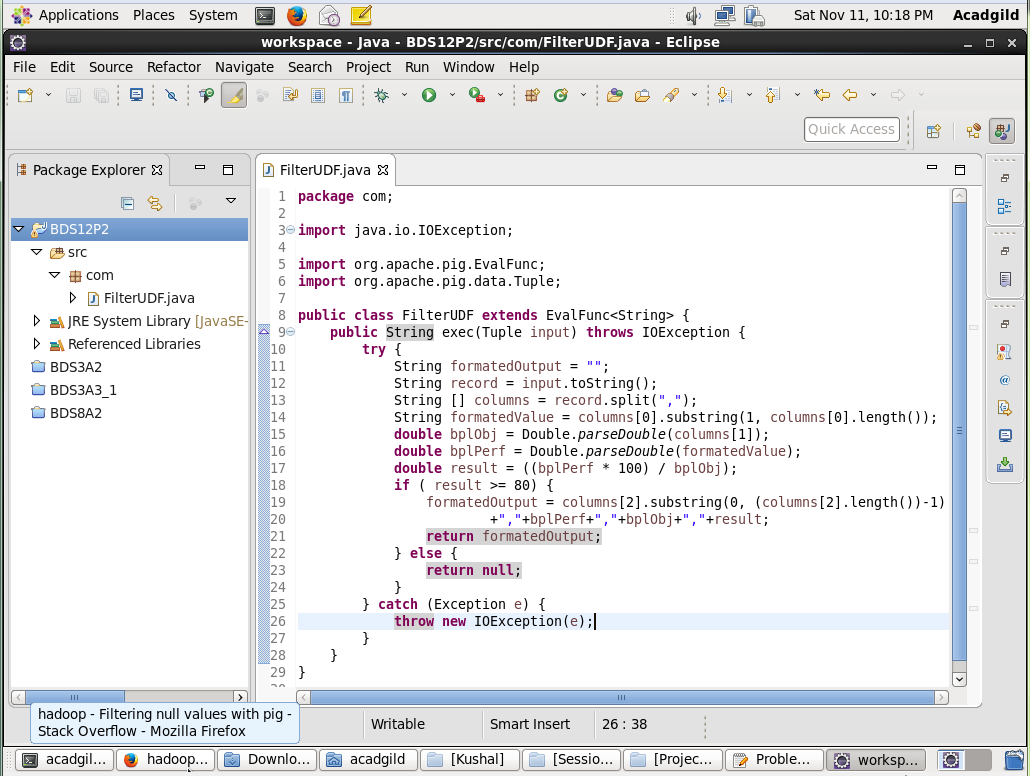


* Log on to MySQL to see the data exported

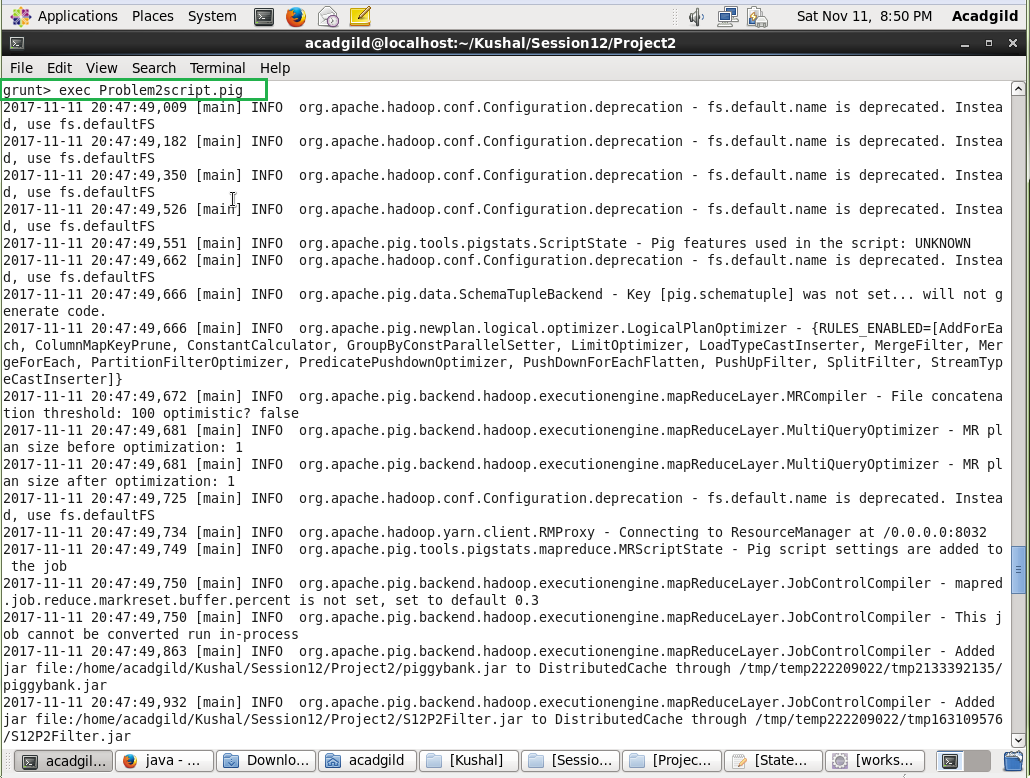


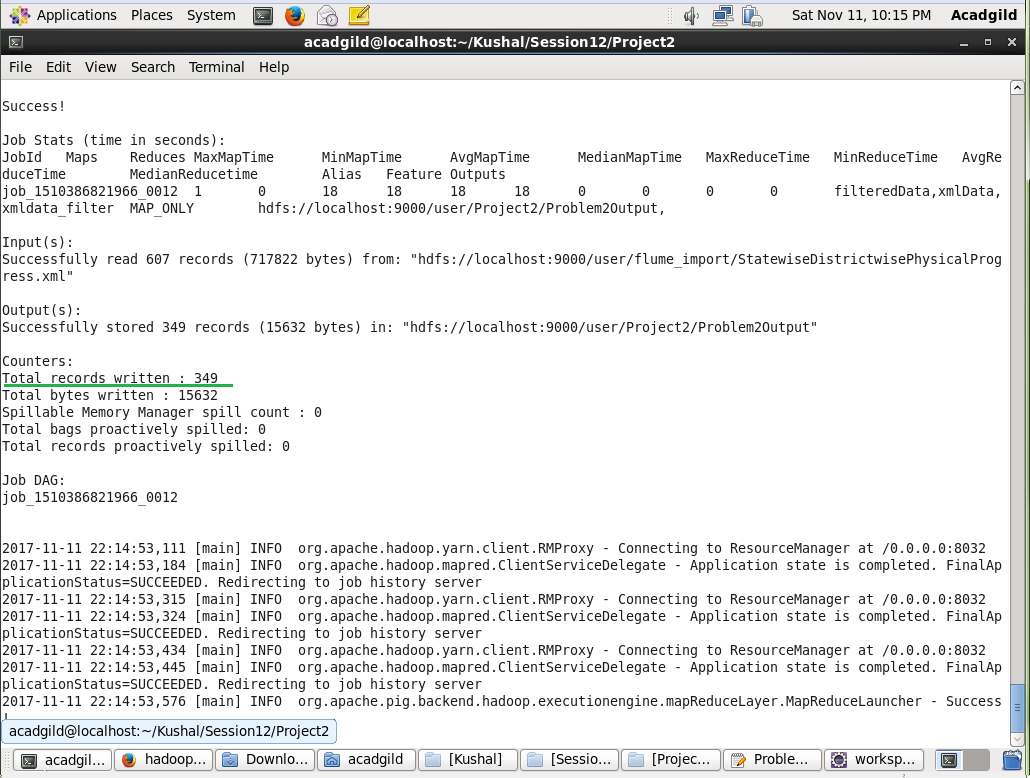
**Problem 2:** Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards. Export the results to MySQL using Sqoop.

* Created a Pig UDF jar to filter.

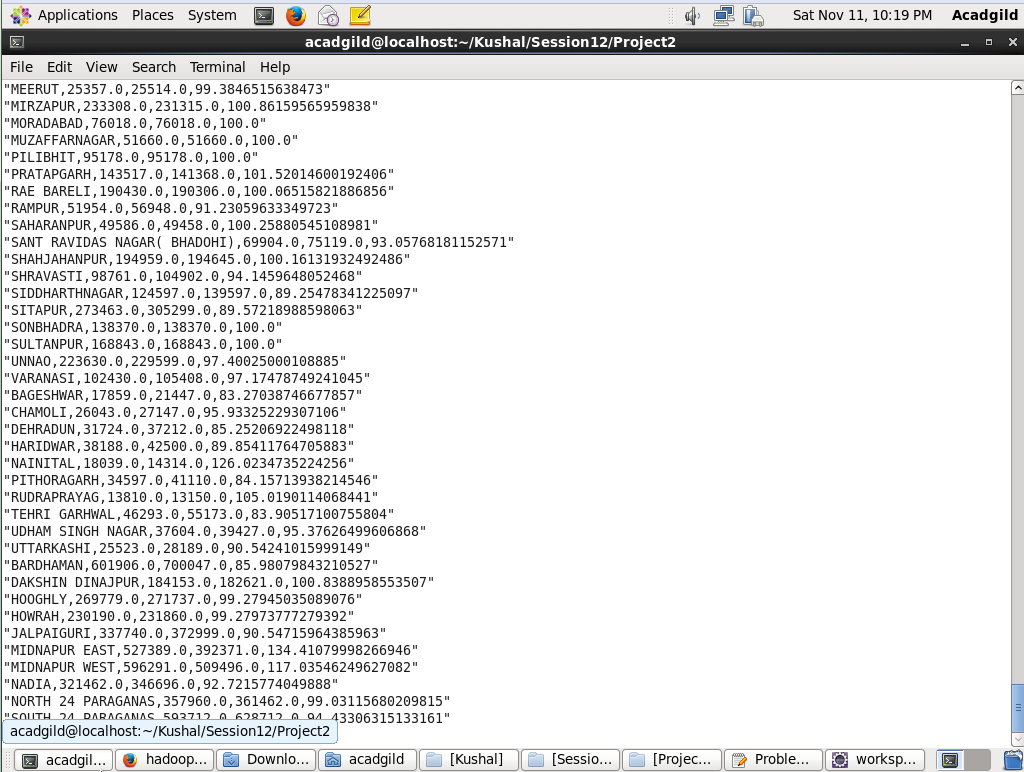


* Running the Problem2script.pig script to get the required Output

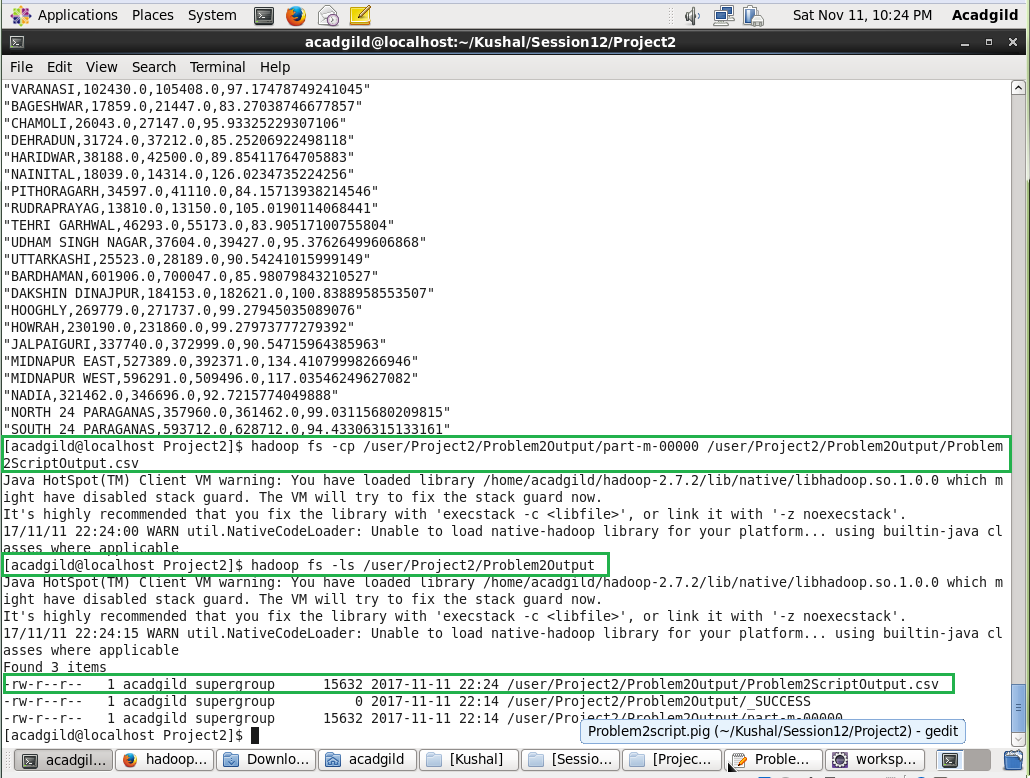




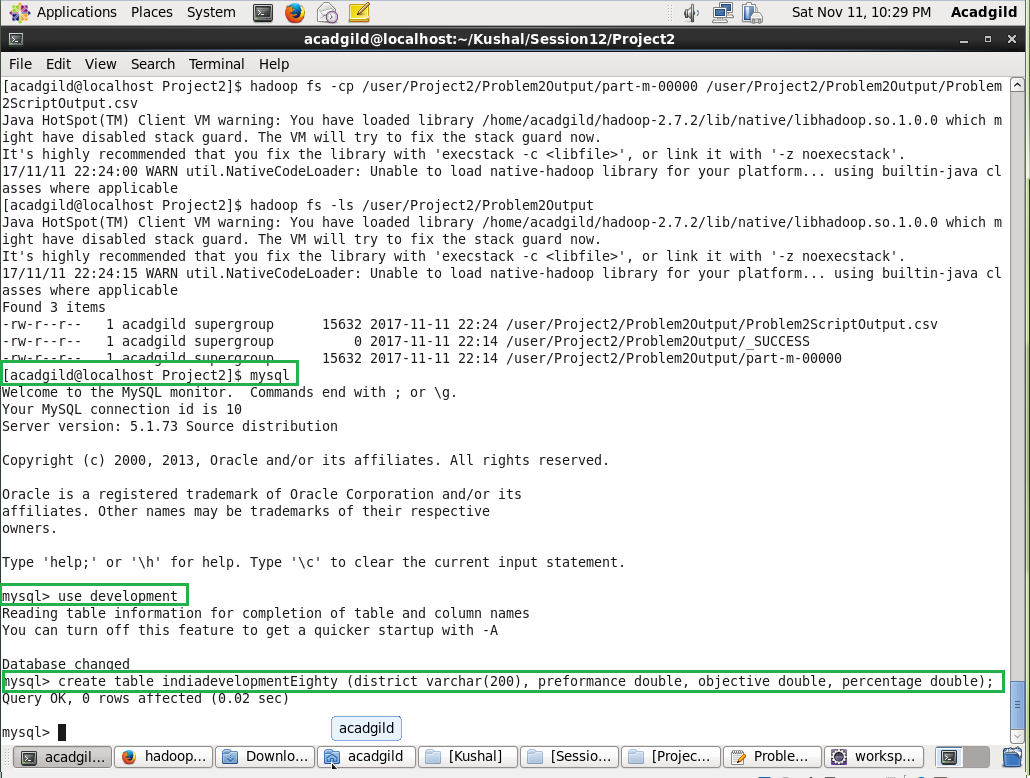
* Output from HDFS



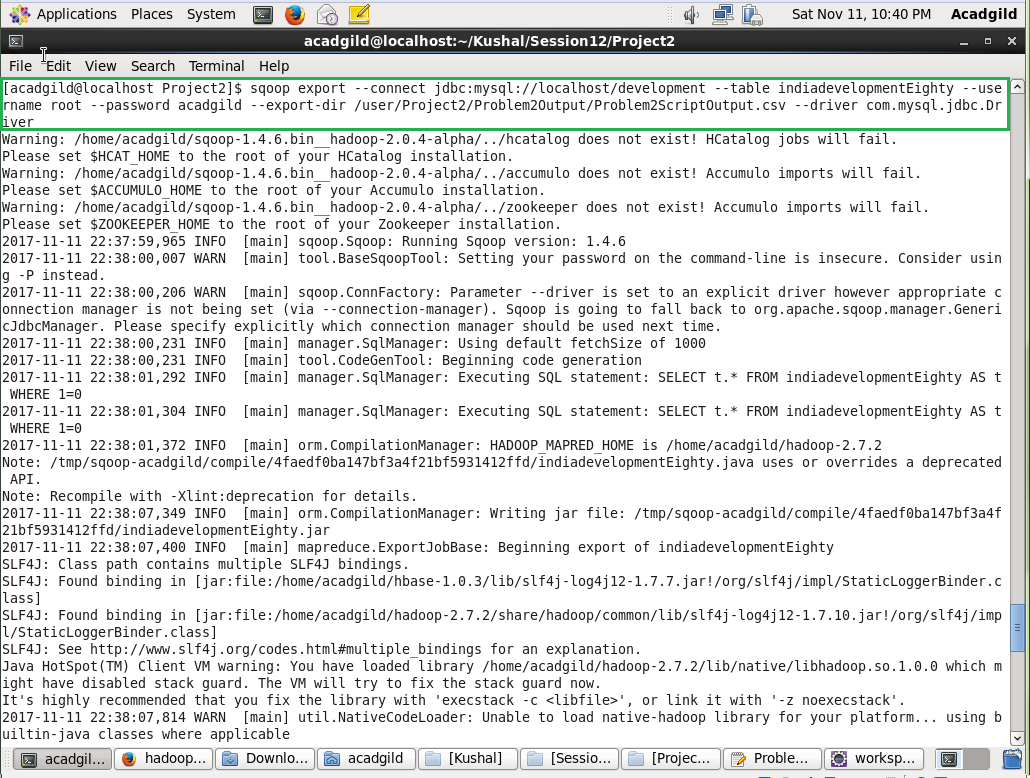
* Convert the output file to .csv to import to MySQL

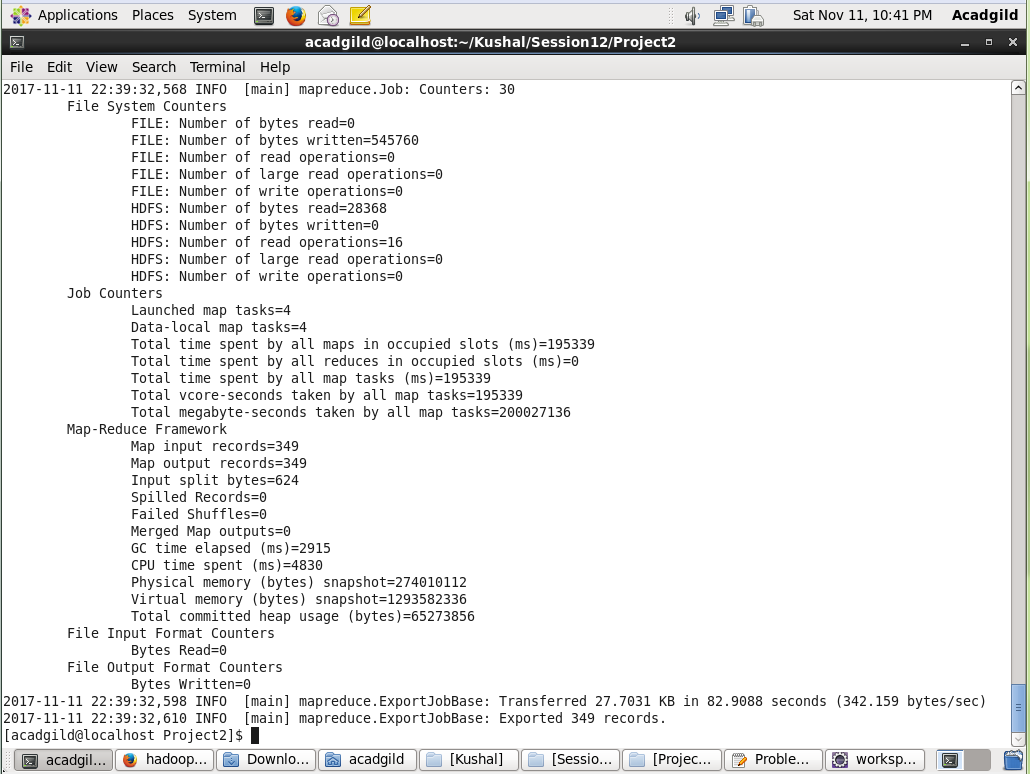


* Logon to MySQL and create a table for storing the .csv data



* Running sqoop command to export .csv data to MySQL





* Log on to MySQL to see the data exported

