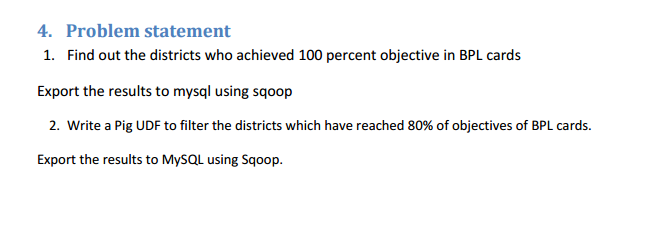
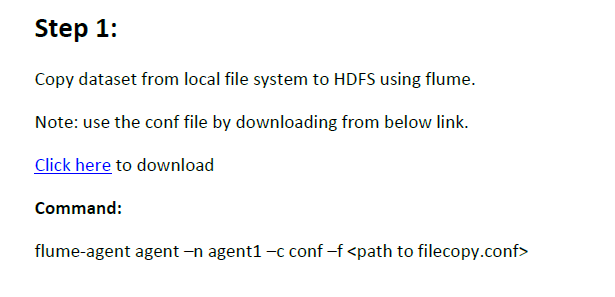
**MAJOR PROJECT**

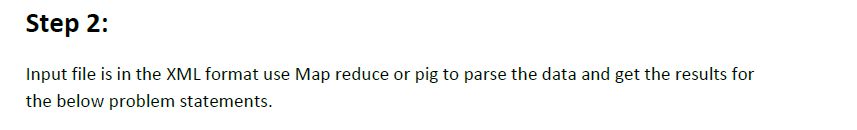
**Problem Statement:**

****

**Step-I**

****

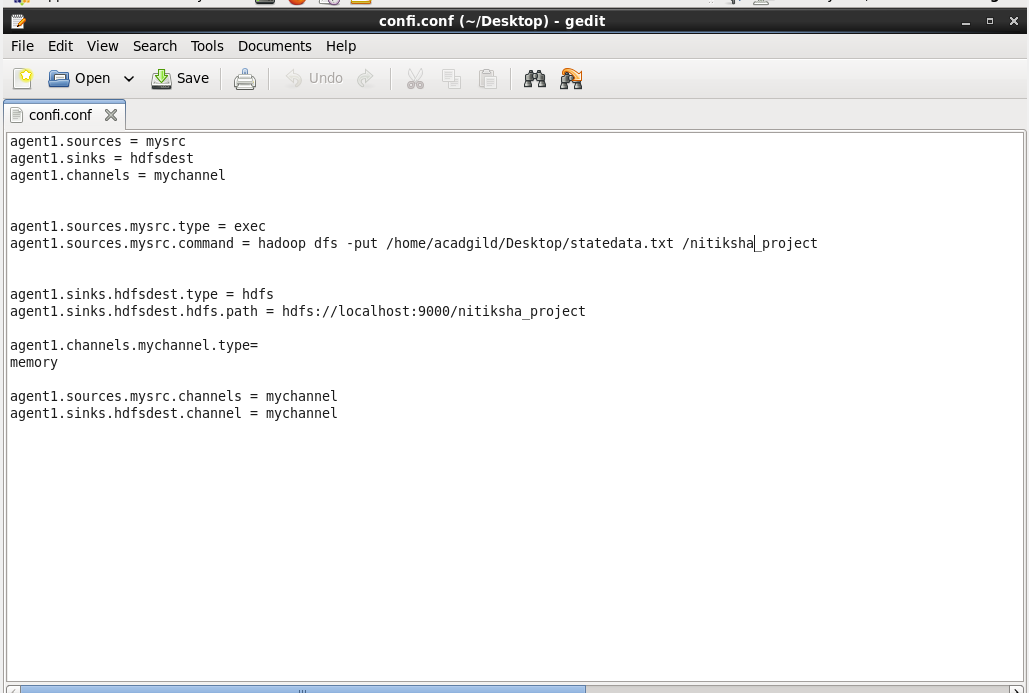
**Step-II**

****

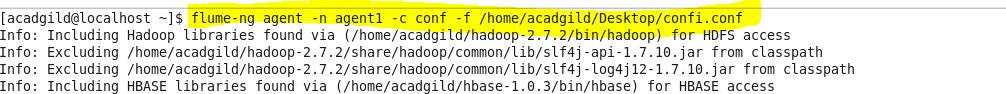
**Solution-**

**Copying dataset from local file system to HDFS using Flume**

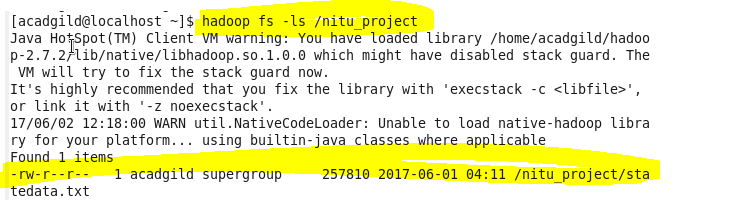
Create conf file and create directory in Hadoop to store local file



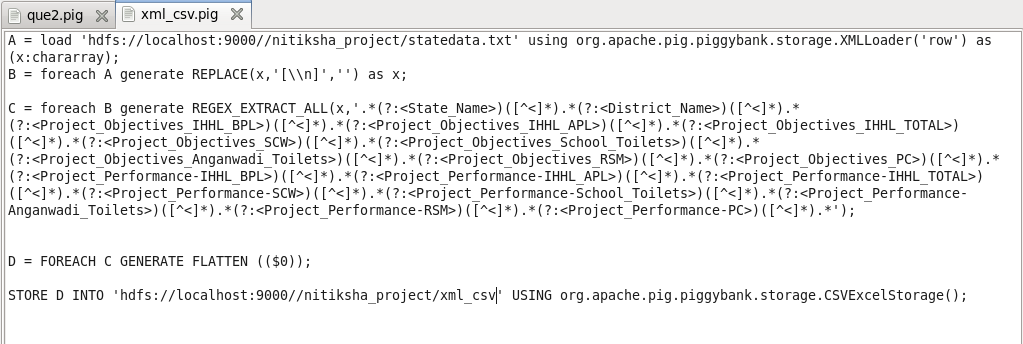
Load and run flume command



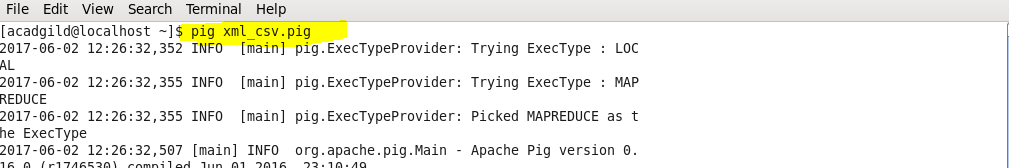
File loaded into HDFS



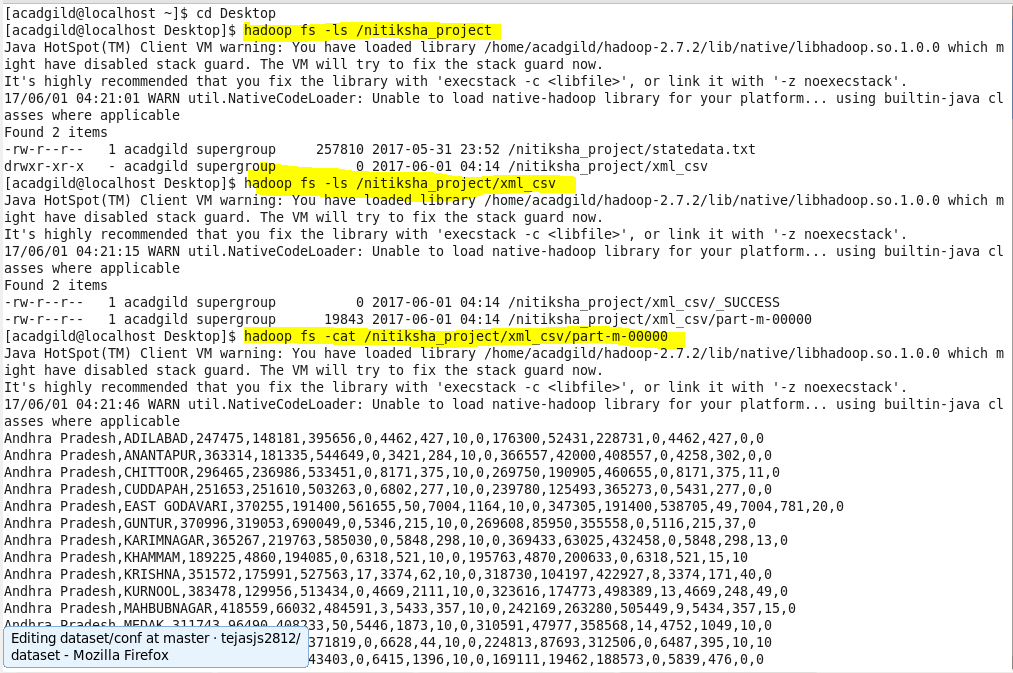
Covert xml to csv, named convert. Pig



Load the convert. Pig



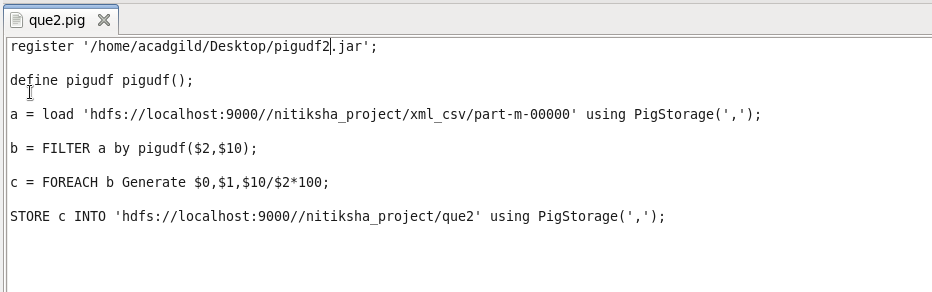
File created into HDFS



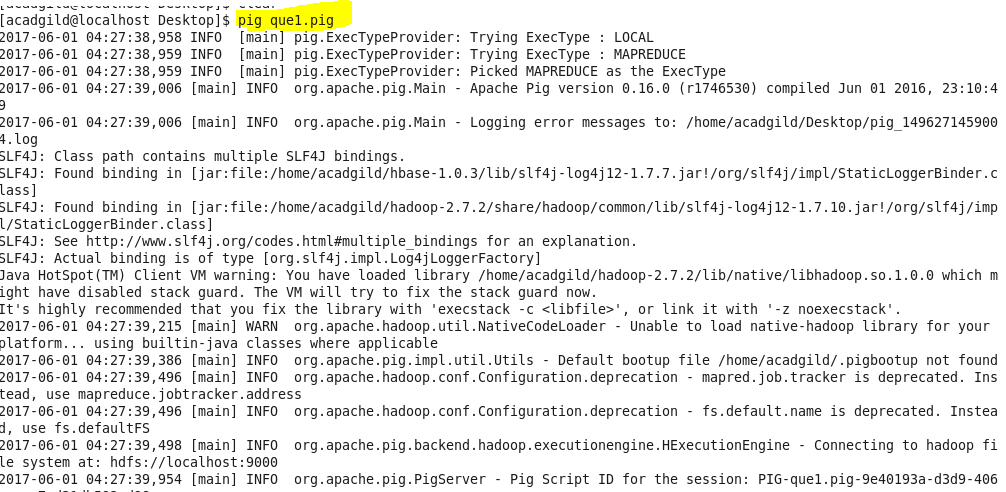
**Que.1) Find out the districts who achieved 100 percent objective in BPL cards**

**Export the results to MySQL using sqoop.**

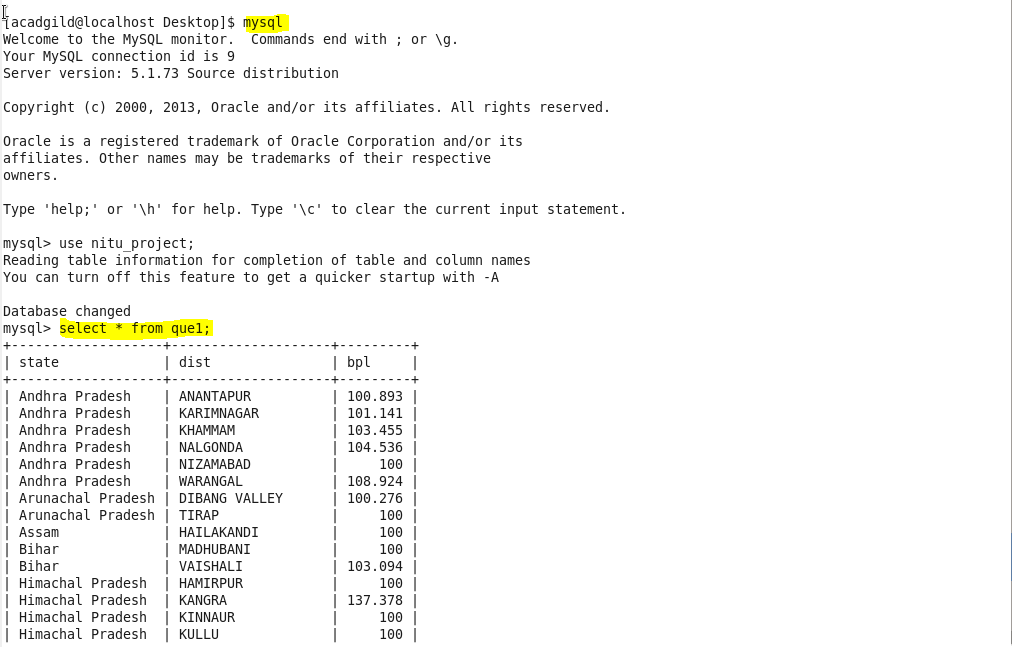
**Create pig script named project-1.pig-**

****

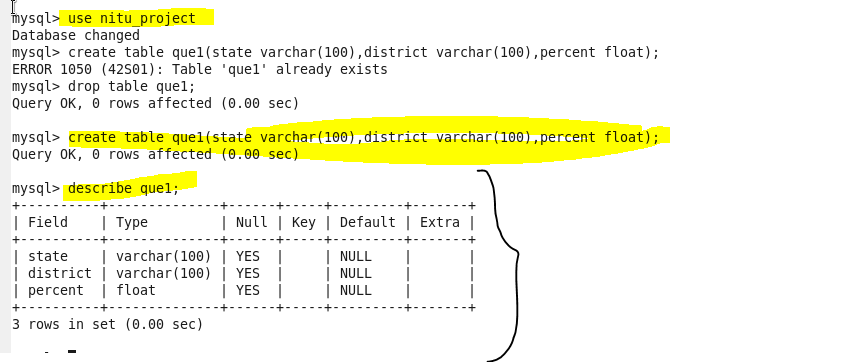
**Run project-1.pig**



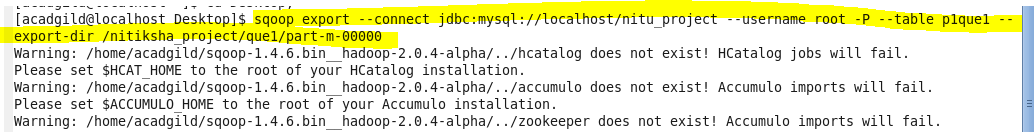
**Display result stored after running pig script-**



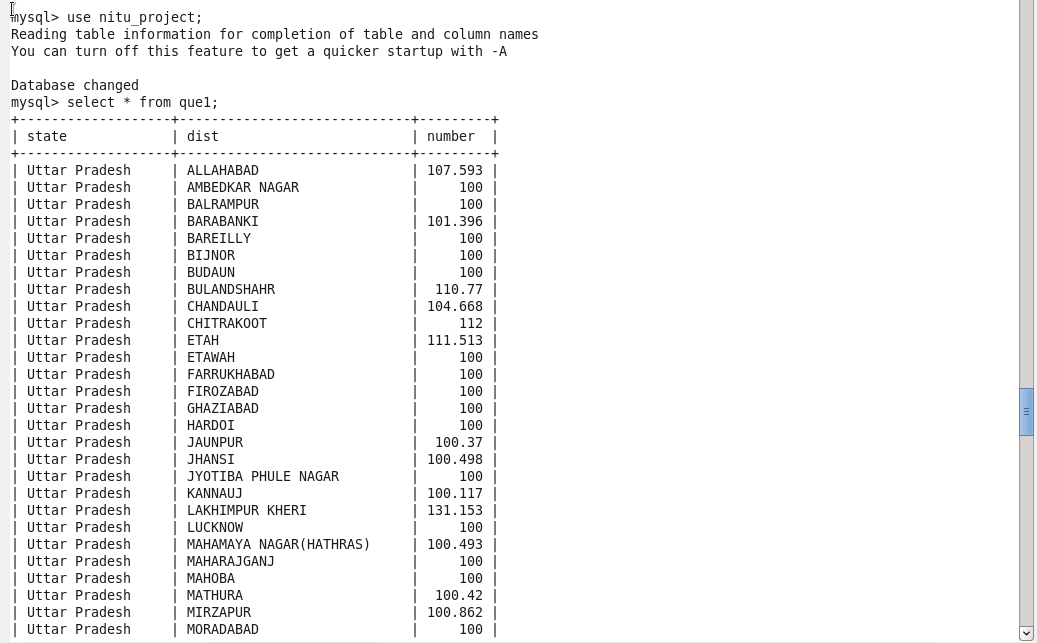
**Create a database and table into mysql and export the result from hdfs to mysql using sqoop-**

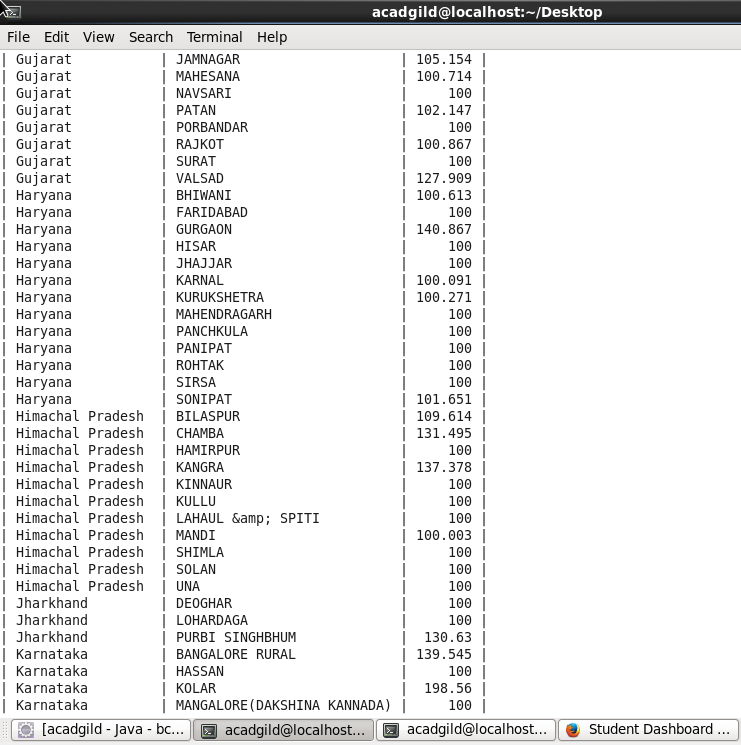


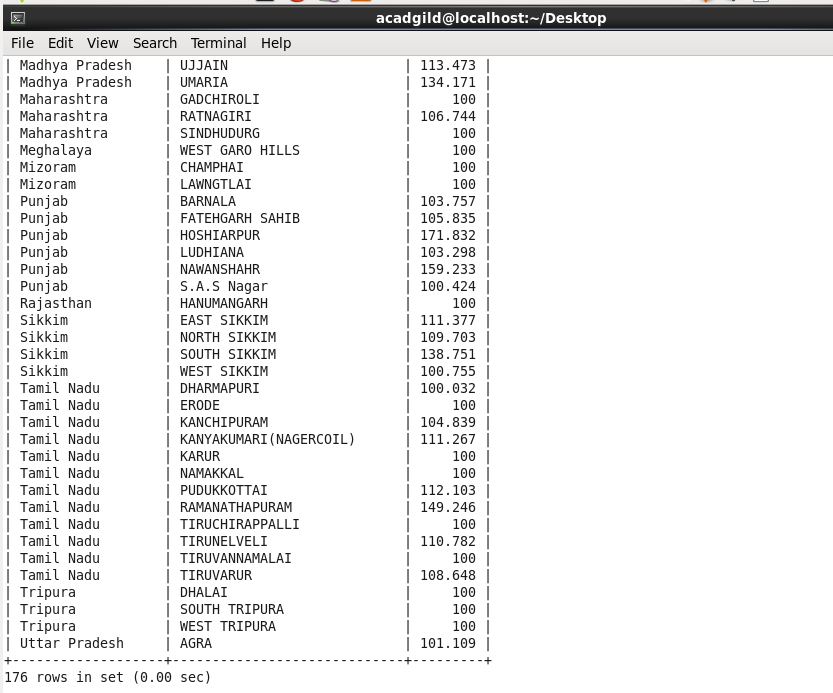
**Export using sqoop**



**Display the result in mysql** –





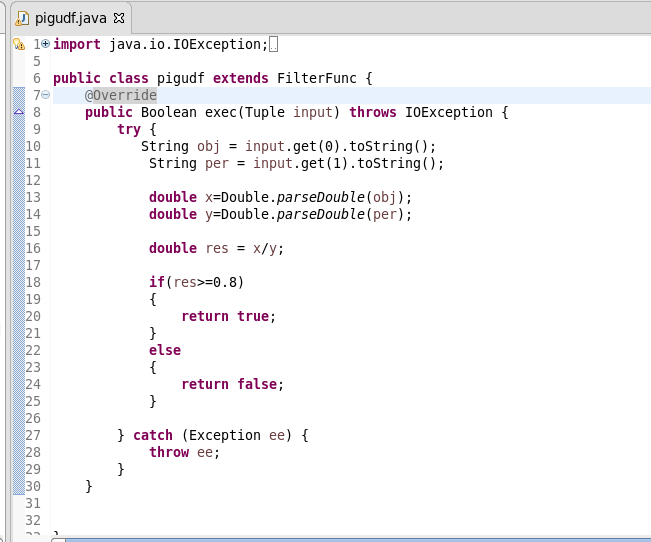


176 ROWS WERE EXPORTED

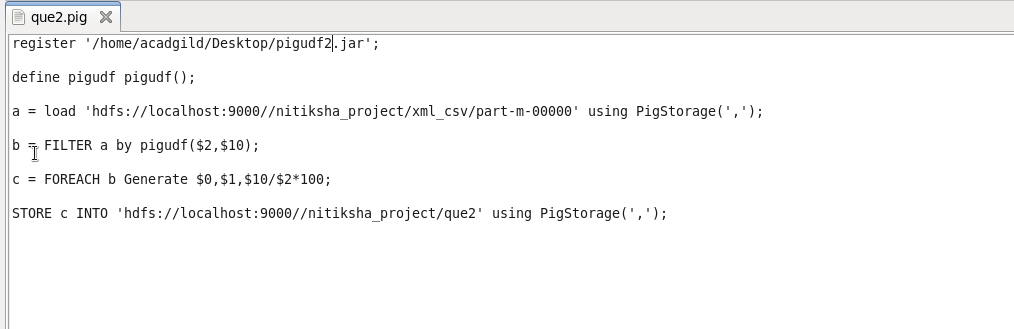
**Que.2) Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards.**

**Export the results to MySQL using Sqoop**.

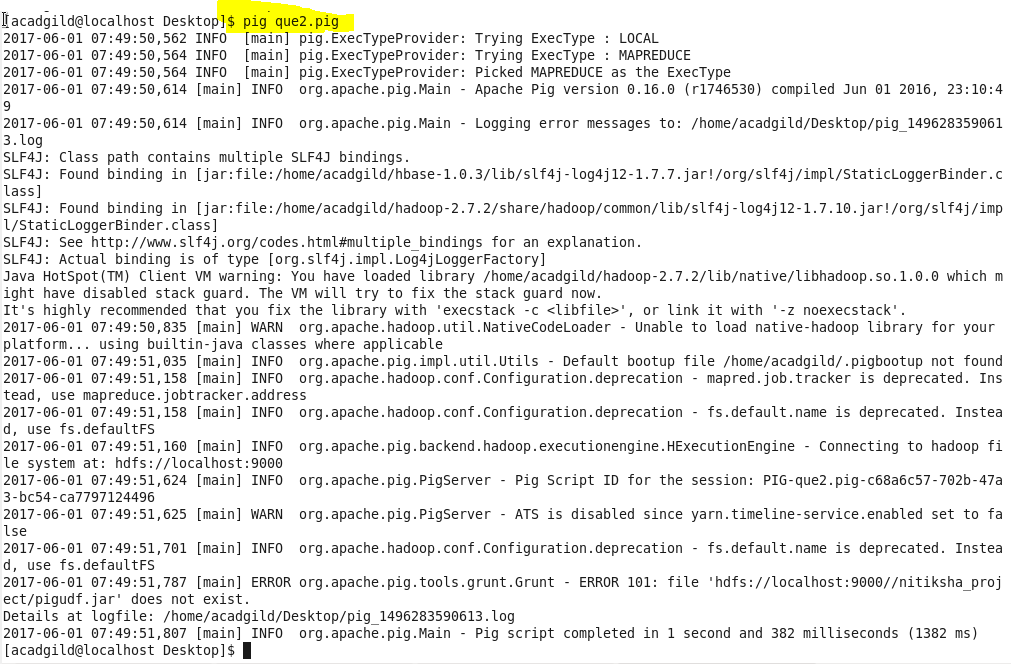
**Created UDF to filter out the result depending on the condition, where the BPL objective is 80% reached, create jar file**

****

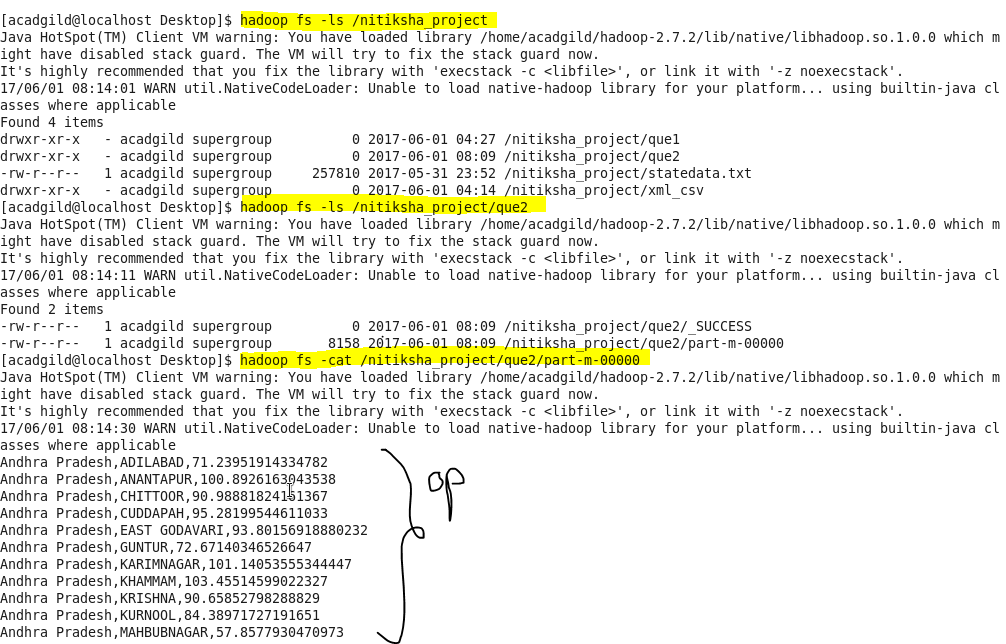
**Load the jar file in pig script , named project-2.pig**

****

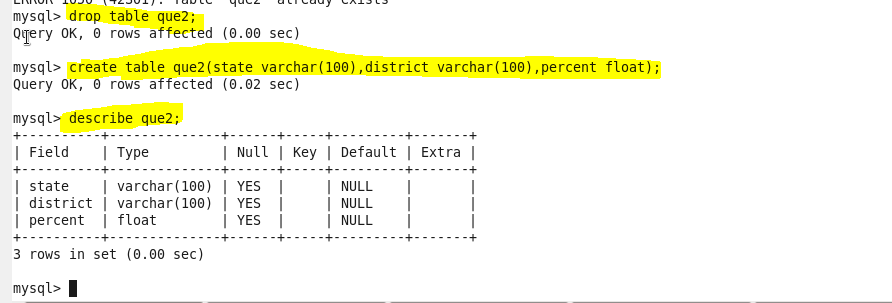
**Run the pig script project-2.pig and store result on hdfs**

****

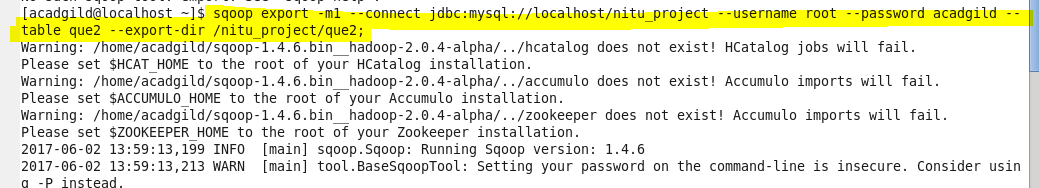
**Display result-**

****

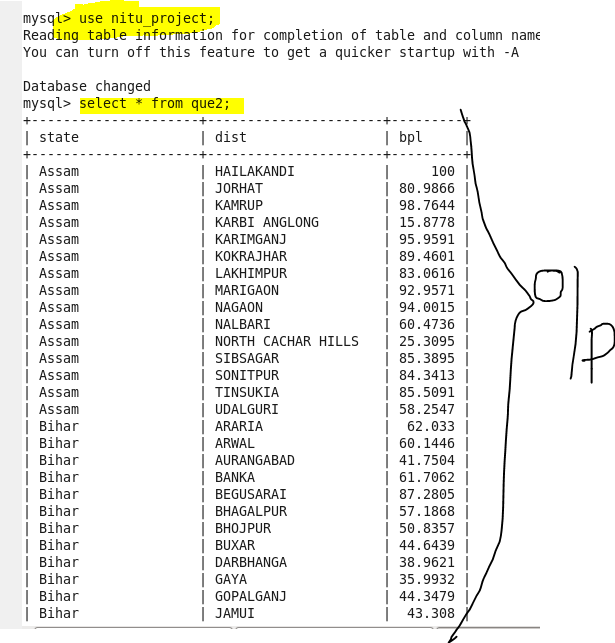
**Create a table in mysql and export the result of project-2 into the table using sqoop-**

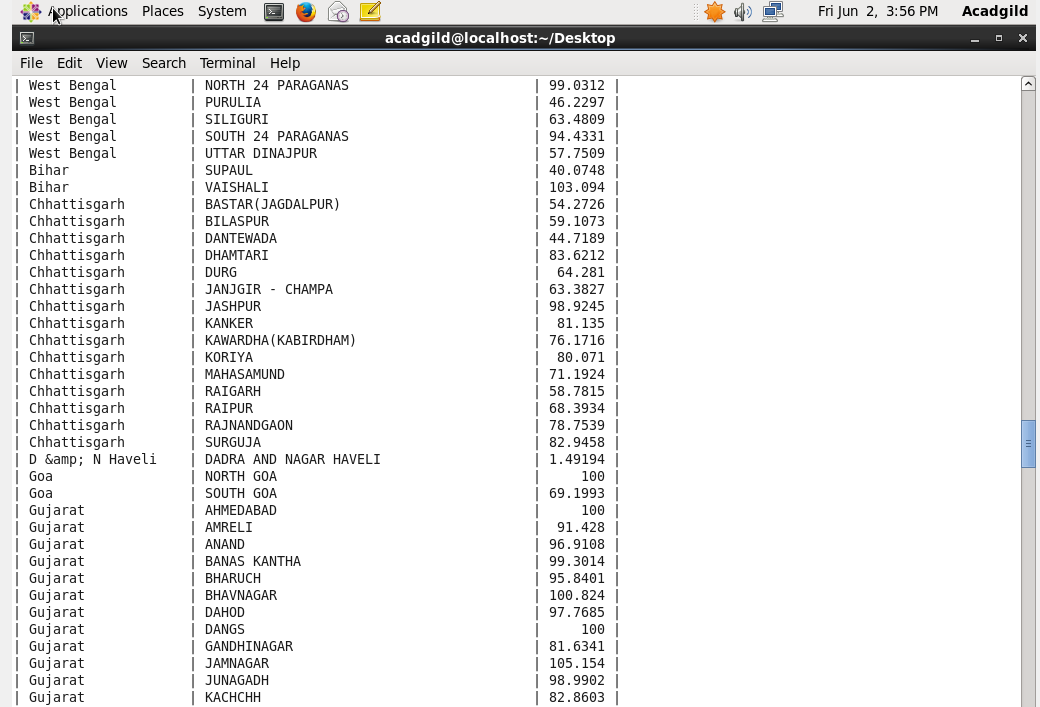
****

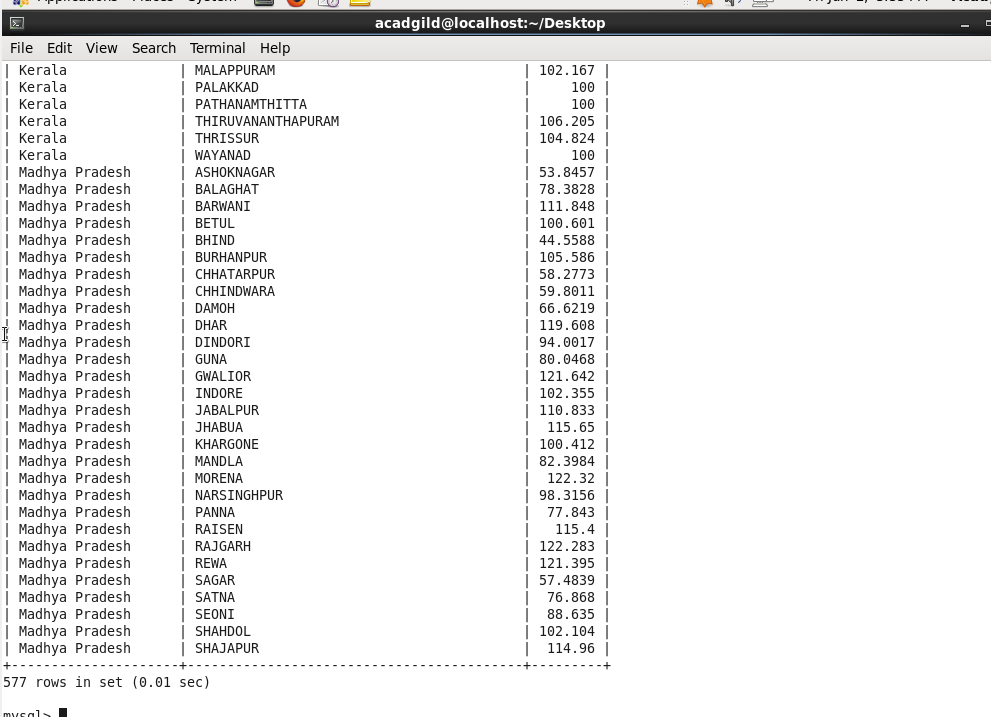
**Sqoop export**

****

**Display result in mysql-**

****

****

****

**577 ROWS WERE EXPORTED IN RESULT**