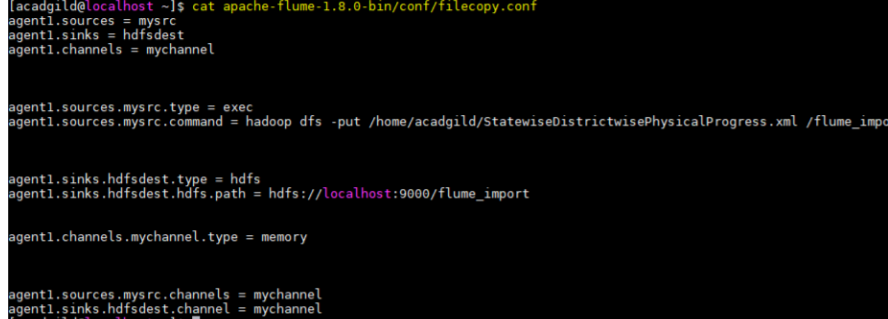
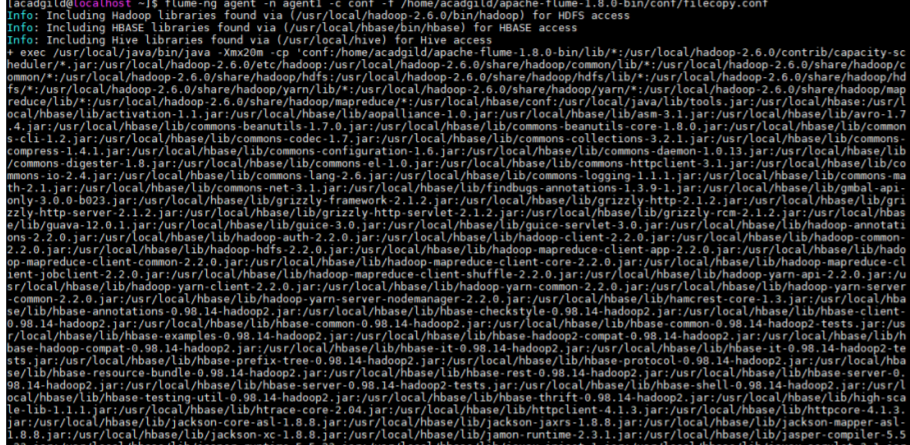
Place the flume config file ‘filecopy.conf’ at location /home/acadgild/apache-flume-1.8.0- bin/conf filecopy.conf



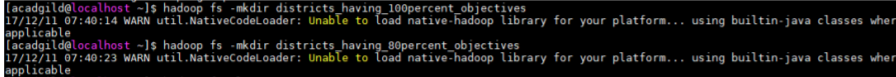
Flume agent



hadoop fs –ls /flume\_import

hadoop fs -mkdir districts\_having\_100percent\_objectives

hadoop fs -mkdir districts\_having\_80percent\_objectives



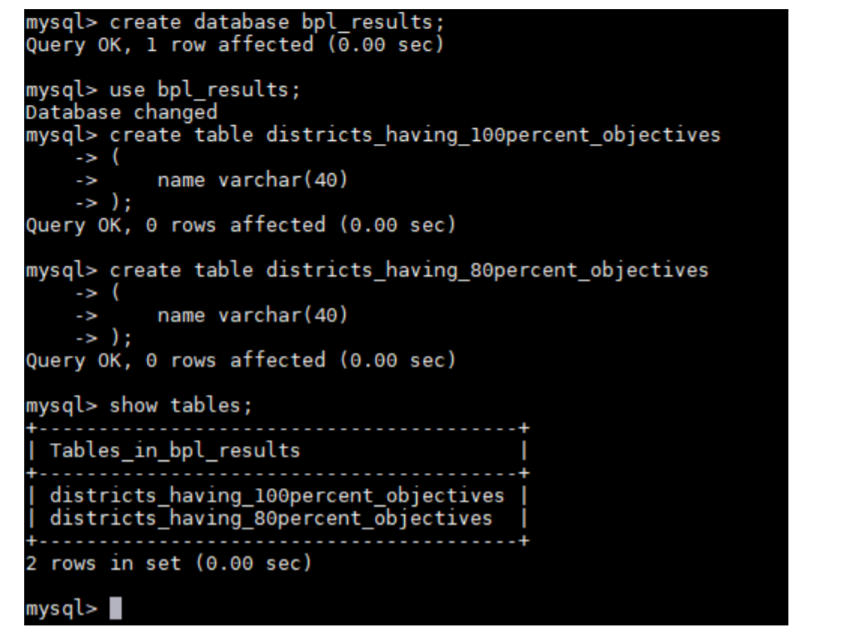
Mysql :create database and tables

create database bpl\_results;

use bpl\_results;

create table districts\_having\_100percent\_objectives ( name varchar(40) );

create table districts\_having\_80percent\_objectives ( name varchar(40) );



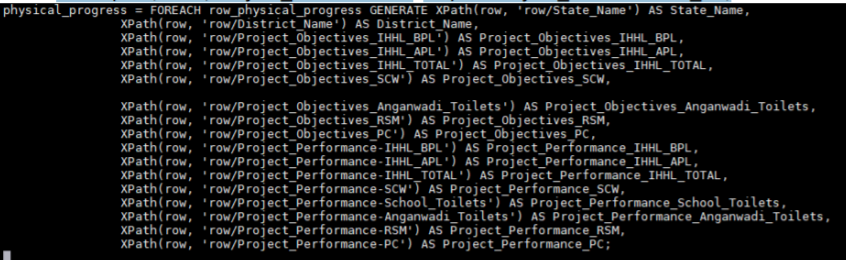
Pig query :

DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath;

row\_physical\_progress = LOAD 'hdfs://localhost:9000/flume\_import' USING org.apache.pig.piggybank.storage.XMLLoader('row') as (row:chararray);

physical\_progress = FOREACH row\_physical\_progress GENERATE XPath(row, 'row/State\_Name') AS State\_Name, XPath(row, 'row/District\_Name') AS District\_Name, XPath(row, 'row/Project\_Objectives\_IHHL\_BPL') AS Project\_Objectives\_IHHL\_BPL, XPath(row, 'row/Project\_Objectives\_IHHL\_APL') AS Project\_Objectives\_IHHL\_APL, XPath(row, 'row/Project\_Objectives\_IHHL\_TOTAL') AS Project\_Objectives\_IHHL\_TOTAL, XPath(row, 'row/Project\_Objectives\_SCW') AS Project\_Objectives\_SCW, XPath(row, 'row/Project\_Objectives\_Anganwadi\_Toilets') AS Project\_Objectives\_Anganwadi\_Toilets, XPath(row, 'row/Project\_Objectives\_RSM') AS Project\_Objectives\_RSM, XPath(row, 'row/Project\_Objectives\_PC') AS Project\_Objectives\_PC, XPath(row, 'row/Project\_Performance-IHHL\_BPL') AS Project\_Performance\_IHHL\_BPL, XPath(row, 'row/Project\_Performance-IHHL\_APL') AS Project\_Performance\_IHHL\_APL, XPath(row, 'row/Project\_Performance-IHHL\_TOTAL') AS Project\_Performance\_IHHL\_TOTAL, XPath(row, 'row/Project\_Performance-SCW') AS Project\_Performance\_SCW, XPath(row, 'row/Project\_Performance-School\_Toilets') AS Project\_Performance\_School\_Toilets, XPath(row, 'row/Project\_Performance-Anganwadi\_Toilets') AS Project\_Performance\_Anganwadi\_Toilets, XPath(row, 'row/Project\_Performance-RSM') AS Project\_Performance\_RSM, XPath(row, 'row/Project\_Performance-PC') AS Project\_Performance\_PC;

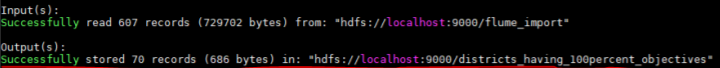




physical\_progress\_100\_percent\_bpl = FILTER physical\_progress BY Project\_Objectives\_IHHL\_BPL == Project\_Performance\_IHHL\_BPL;

district\_100\_percent\_bpl = FOREACH physical\_progress\_100\_percent\_bpl GENERATE District\_Name; STORE district\_100\_percent\_bpl INTO 'hdfs://localhost:9000/districts\_having\_100percent\_objectives';

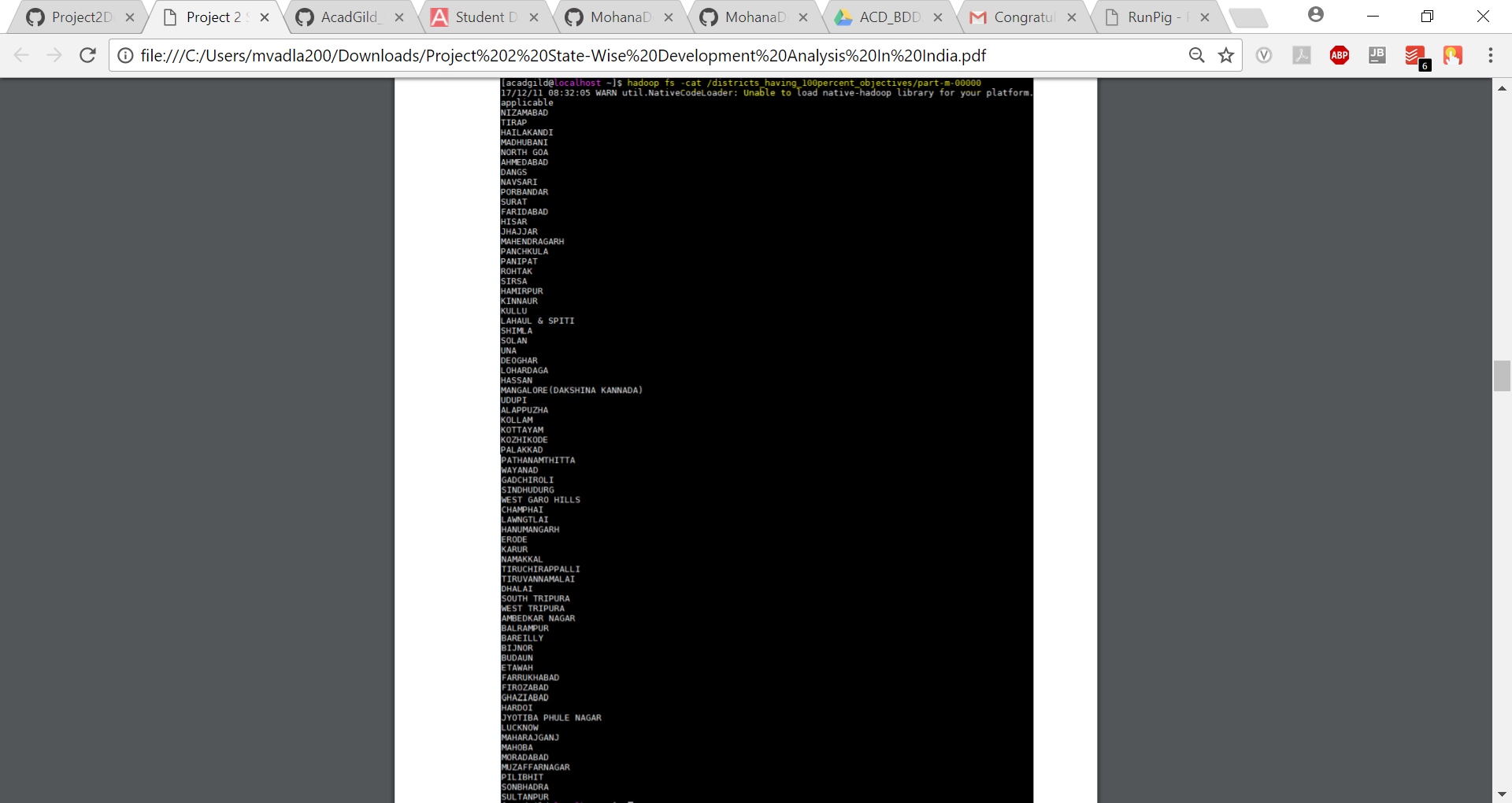




Verify results:

hadoop fs -ls /districts\_having\_100percent\_objectives

hadoop fs -ls /districts\_having\_100percent\_objectives/part-m-00000

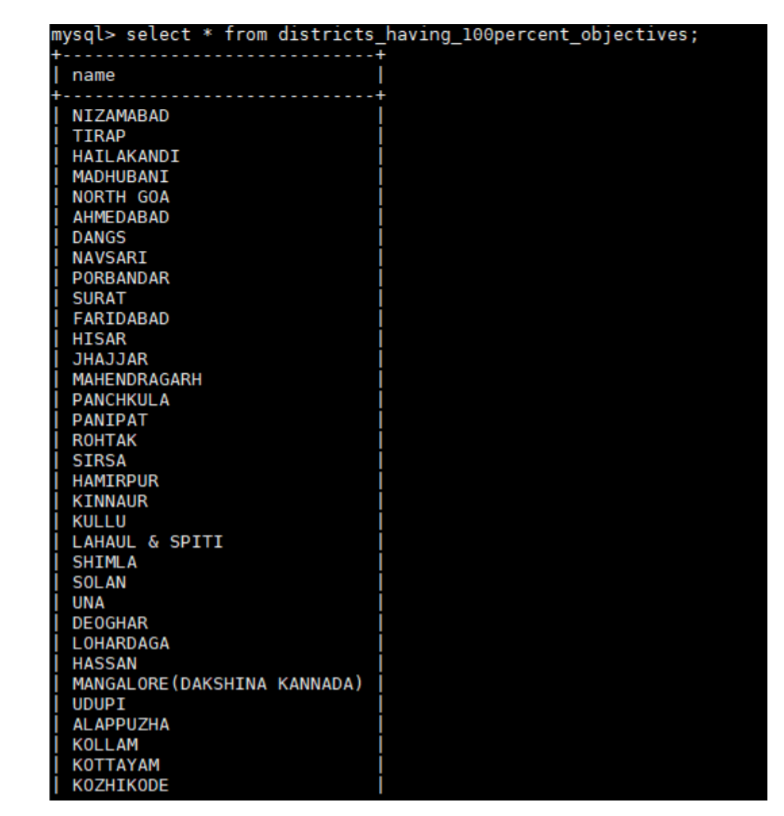


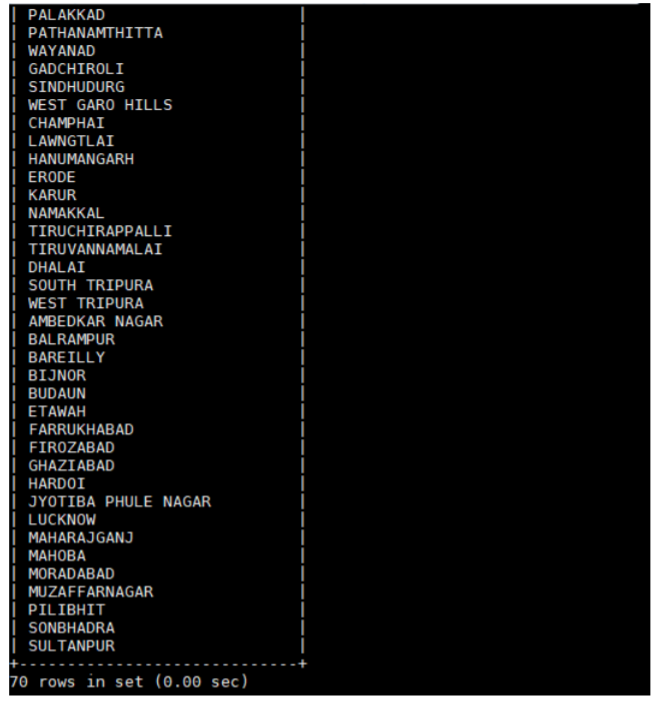
The following sqoop command is used to export data from HDFS folder



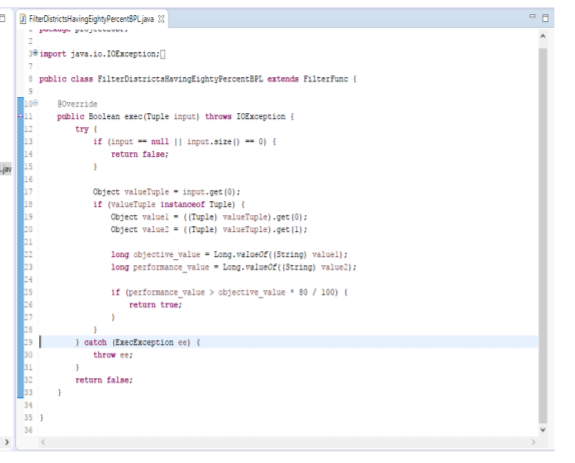
Verify results in mysql

select \* from districts\_having\_100percent\_objectives;





Create a Java project Project2 and Write a Java class FilterDistrictsHavingEightyPercentBPL in eclipse which will filter those tuples for which 80 percent objective in BPL cards are achieved. The logic put in exec method is value of Project\_Performance\_IHHL\_BPL is equal to more than 80% of Project\_Objectives\_IHHL\_BPL.



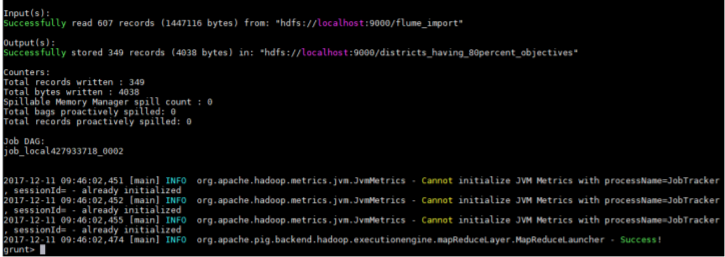
REGISTER /home/acadgild/pig/Project2.jar;

physical\_progress\_80\_percent\_bpl = FILTER physical\_progress BY project2UDF.FilterDistrictsHavingEightyPercentBPL(TOTUPLE(Project\_Objectives\_IHHL\_BPL, Project\_Performance\_IHHL\_BPL));

district\_80\_percent\_bpl = FOREACH physical\_progress\_80\_percent\_bpl GENERATE District\_Name;

STORE district\_80\_percent\_bpl INTO 'hdfs://localhost:9000/districts\_having\_80percent\_objectives';

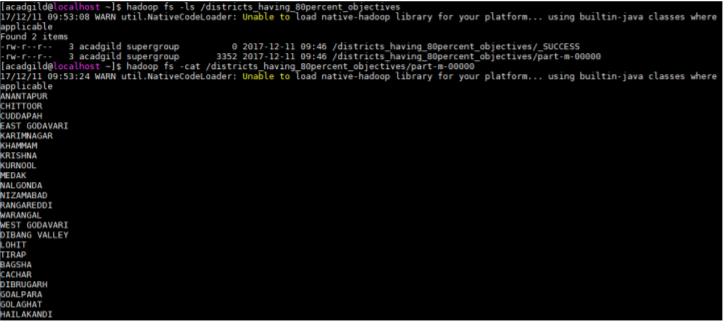




Verify results:

hadoop fs -ls /districts\_having\_80percent\_objectives

hadoop fs –ls /districts\_having\_80percent\_objectives/part-m-00000



Use sqoop command to export data from HDFS into mysql table districts\_having\_80percent\_objectives in database bpl\_result

sqoop export --connect jdbc:mysql://localhost/bpl\_results --username 'root' --table 'districts\_having\_80percent\_objectives' --export-dir '/districts\_having\_80percent\_objectives' -- input-fields-terminated-by ',' -m 1 --columns name



Mysql results:

select \* from districts\_having\_80percent\_objectives

