statement 2: Write a Pig UDF to filter the districts who have reached 80% of objectives of BPL cards.

Export the results to mysql using sqoop.

Create a Pig UDF Filter function to calculate districts that have met 80% or more of the BPL card Objectives.

Convert it into a Jar file to implement as a UDF filter in Pig Script

To create the Jar File, Open eclipse, Create a new java Project: CalcPC.java and add external jar Files from

---

1) /usr/local/pig

2) /usr/local/hadoop-2.6.0/share/hadoop/common/lib

3) /usr/local/hadoop-2.6.0/share/hadoop/common

4) /usr/local/pig/lib/hadoop1-runtime

CalcPC.java

```
import java.io.IOException;

import org.apache.pig.FilterFunc;

import org.apache.pig.data.Tuple;



public class CalcPc extends FilterFunc {

        public Boolean exec (Tuple input) throws IOException

        {

                try

                {

                        int value1 = Integer.parseInt((String) input.get(0));

                        int value2 = Integer.parseInt((String) input.get(1));
```

```
                if (value1 == 0 || value2 == 0 )

                {

                        System.out.println("zero values");

                        System.exit(1);

                }

                return ((value1/value2)>=0.8);

        }

        catch (Exception e)

        {

                System.out.println("something wrong"+e.getMessage());

        }

        return null;

}
```

Export the java class to state.jar

state.pig

```
REGISTER /home/acadgild/state.jar;

REGISTER /usr/local/pig/lib/piggybank.jar;



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



Data =  LOAD 'statewisedevelopment/state.xml' using
org.apache.pig.piggybank.storage.XMLLoader('row') as (x:chararray);

StateDet = FOREACH Data GENERATE XPath(x, 'row/State_Name') AS statename, XPath(x,
'row/District_Name') AS disname,XPath(x, 'row/Project_Objectives_IHHL_BPL') AS BPL,XPath(x,
'row/Project_Objectives_IHHL_TOTAL') AS total ;

FilteredData = FILTER StateDet BY CalcPc(BPL,total);

STORE FilteredData INTO '/home/acadgild/statepc' USING PigStorage(',');
```

The Output is now present in the directory statepc

[acadgild@localhost ~]$ ls /home/acadgild/statepc

part-m-00000 _SUCCESS

Contents of the output file

[acadgild@localhost ~]$ cat /home/acadgild/statepc/p*

Arunachal Pradesh,ANJAW,3232,3232

Arunachal Pradesh,DIBANG VALLEY,1085,1085

Arunachal Pradesh,KURUNG KUMEY,22036,22036

Arunachal Pradesh,LOHIT,8800,8800

Arunachal Pradesh,WEST SIANG,11472,11472

Bihar,BANKA,82439,82439

D & N Haveli,DADRA AND NAGAR HAVELI,2480,2480

Goa,NORTH GOA,15000,15000

Jammu & Kashmir,KARGIL,8475,8475

Jammu & Kashmir,KISHTWAR,22318,22318

Jammu & Kashmir,LEH (LADAKH),6090,6090

Jammu & Kashmir,REASI ,21500,21500

Jammu & Kashmir,SAMBA ,9849,9849

Jammu & Kashmir,SHOPIAN ,10196,10196

Copy the Directory to HDFS to prepare for Sqoop export

[acadgild@localhost ~]$ hadoop fs -put statepc /user/acadgild

Make sure the copy is successful

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild

17/03/09 00:17:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 7 items

drwxr-xr-x   - acadgild supergroup   0 2017-03-08 23:45 /user/acadgild/statepc

[acadgild@localhost ~]$

Open Terminal2,

1) start all Hadoop Daemons using start-all.sh

2)Make sure they are up using

[acadgild@localhost ~]$ jps

16884 Jps

3382 SecondaryNameNode

3560 ResourceManager

3118 NameNode

3663 NodeManager

3215 DataNode

3) Run MySQL

sudo service mysqld start

mysql -u root

mysql>use state;

create a table similar to the Output structure

mysql>create table state80percent (State varchar(20), district varchar(50), BPL int, total int);

mysql>show tables;

+------------------+

| Tables_in_state |

+------------------+

| BPLObjectivesMet |

| state80percent |

+------------------+

2 rows in set (0.00 sec)

In terminal1, run the sqoop export command to transfer output from hdfs to MySQL

---

sqoop export --connect jdbc:mysql://localhost/state --username 'root' -P --table state80percent --export-dir '/user/acadgild/statepc/part-m-00000' --input-fields-terminated-by ',' -m 1

In terminal2, check if the table has been populated with data

mysql> select * from state80percent;

+------------------+--+-------+-------+

| State | district | BPL | total |

+------------------+--+-------+-------+

| Arunachal Pradesh | ANJAW | 3232 | 3232 |

| Arunachal Pradesh | DIBANG VALLEY | 1085 | 1085 |

| Arunachal Pradesh | KURUNG KUMEY | 22036 | 22036 |

| Arunachal Pradesh | LOHIT | 8800 | 8800 |

| Arunachal Pradesh | WEST SIANG | 11472 | 11472 |

| Bihar | BANKA | 82439 | 82439 |

| D & N Haveli | DADRA AND NAGAR HAVELI | 2480 | 2480 |

| Goa | NORTH GOA | 15000 | 15000 |

| Jammu & Kashmir | KARGIL | 8475 | 8475 |

| Jammu & Kashmir | KISHTWAR | 22318 | 22318 |

| Jammu & Kashmir | LEH (LADAKH) | 6090 | 6090 |

| Jammu & Kashmir | REASI | 21500 | 21500 |

| Jammu & Kashmir | SAMBA | 9849 | 9849 |

| Jammu & Kashmir | SHOPIAN | 10196 | 10196 |

```
+------------------+----------------------------------------+-------+-------+
```

14 rows in set (0.01 sec)