## PROBLEM2

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

Export the results to mysql using sqoop.

--defining xml-xpath for importing datas

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: Calc. java and add external jar Files

## Calc.java

```
package pigudf;
import java.io.IOException;
import org.apache.pig.FilterFunc;
import org.apache.pig.data.Tuple;
public class Calc 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;
      }
}
problem2.pig
--register piggybank lib
REGISTER /home/acadgild/piggybank-0.15.0.jar
REGISTER /home/acadgild/stateWise.jar;
```

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

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

--reading proper values for computation from xml file

State = 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;

--creating temporary function

DEFINE func pigudf.Calc;

--filtering as per logic

Obj = FILTER State BY func(BPL,total);

--store result

STORE Obj INTO '/vishnu/pigudf' USING PigStorage(',');

\_\_\_\_\_

The Output is now present in the directory /vishnu/pigudf

[acadgild@localhost ~]\$ hadoop fs -cat /vishnu/pigudf/p\*

17/11/07 18:22:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

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

Kerala,KANNUR,34121,34121 Manipur,CHANDEL,17610,17610 Nagaland,LONGLENG,6438,6438 Nagaland, TUENSANG, 13027, 13027 Nagaland, ZUNHEBOTO, 20570, 20570 Puducherry, PONDICHERRY, 18000, 18000 Punjab, FARIDKOT, 6000, 6000 Punjab, HOSHIARPUR, 11112, 11112 Punjab, MOGA, 37170, 37170 Punjab, MUKTSAR, 33148, 33148

Copy the Directory to HDFS to prepare for Sqoop export

Run MySQL sudo service mysqld start mysql -u root

mysql> use state;

Reading table information for completion of table and column names You can turn off this feature to get a quicker startup with -A

Database changed

mysql> create table state80 (State varchar(20), district varchar(50), BPL int, total int); Query OK, 0 rows affected (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 state80 --export-dir '/vishnu/pigudf/part-m-00000' --input-fields-terminated-by ',' -m 1

```
mysql> select * from state80; +-----+
| State | district | BPL | total | +-----+
                                   | 3232 | 3232 |
| Arunachal Pradesh | ANJAW
| 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 |
| 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 |
| Kerala
             | KANNUR
                                            | 34121 | 34121 |
| Manipur
                                             | 17610 | 17610 |
               | CHANDEL
Nagaland
                                               | 6438 | 6438 |
               LONGLENG
| Nagaland
                                               | 13027 | 13027 |
               | TUENSANG
| Nagaland
                                               | 20570 | 20570 |
               | ZUNHEBOTO
| Puducherry
               | PONDICHERRY
                                                 | 18000 | 18000 |
                                            | 6000 | 6000 |
| Punjab
              | FARIDKOT
                                               | 11112 | 11112 |
| Punjab
              | HOSHIARPUR
| Punjab
              | MOGA
                                          | 37170 | 37170 |
| Punjab
              | MUKTSAR
                                             | 33148 | 33148 |
24 rows in set (0.00 sec)
```

## Scree shots

```
Input(s):
Successfully read 0 records from: "/pig/state.xml"
Output(s):
Successfully stored 0 records in: "/vishnu/piqudf"
Counters:
Total records written: 0
Total bytes written: 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job 1510047883177 0005
2017-11-07 18:20:27,788 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
necting to ResourceManager at /0.0.0.0:8032
2017-11-07 18:20:27,811 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect:
ing to job history server
2017-11-07 18:20:27,986 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
```

[acadgild@localhost ~]\$ hadoop fs -cat /vishnu/pigudf/p\*

17/11/07 18:22:36 WARN util.NativeCodeLoader: Unable to load native-hadoop libra

ry for your platform... using builtin-java classes where applicable

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

Jammu & Kashmir, SAMBA

Jammu & Kashmir, SHOPIAN

Kerala, KANNUR, 34121, 34121

Manipur, CHANDEL, 17610, 17610

Nagaland, LONGLENG, 6438, 6438

Nagaland, TUENSANG, 13027, 13027

Nagaland, ZUNHEBOTO, 20570, 20570

Puducherry, PONDICHERRY, 18000, 18000

Punjab, FARIDKOT, 6000, 6000

Punjab, HOSHIARPUR, 11112, 11112

Punjab, MOGA, 37170, 37170

Punjab, MUKTSAR, 33148, 33148

[acadgild@localhost ~]\$

,21500,21500

,9849,9849

,10196,10196

nysql> select \* from state80;

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
Kerala	KANNUR	34121	34121
Manipur	CHANDEL	17610	17610
Nagaland	LONGLENG	6438	6438
Nagaland	TUENSANG	13027	13027
Nagaland	ZUNHEBOTO	20570	20570
Puducherry	PONDICHERRY	18000	18000
Punjab	FARIDKOT	6000	6000
Punjab	HOSHIARPUR	11112	11112
Punjab	MOGA	37170	37170
Punjab	MUKTSAR	33148	33148
+	Sh   ANJAW   3232   3232   Sh   DIBANG VALLEY   1085   1085   Sh   KURUNG KUMEY   22036   22036   Sh   LOHIT   8800   8800   Sh   WEST SIANG   11472   11472   1472   8800   82439   8		

24 rows in set (0.00 sec)