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.

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;

--defining xml-xpath for importing datas

```
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

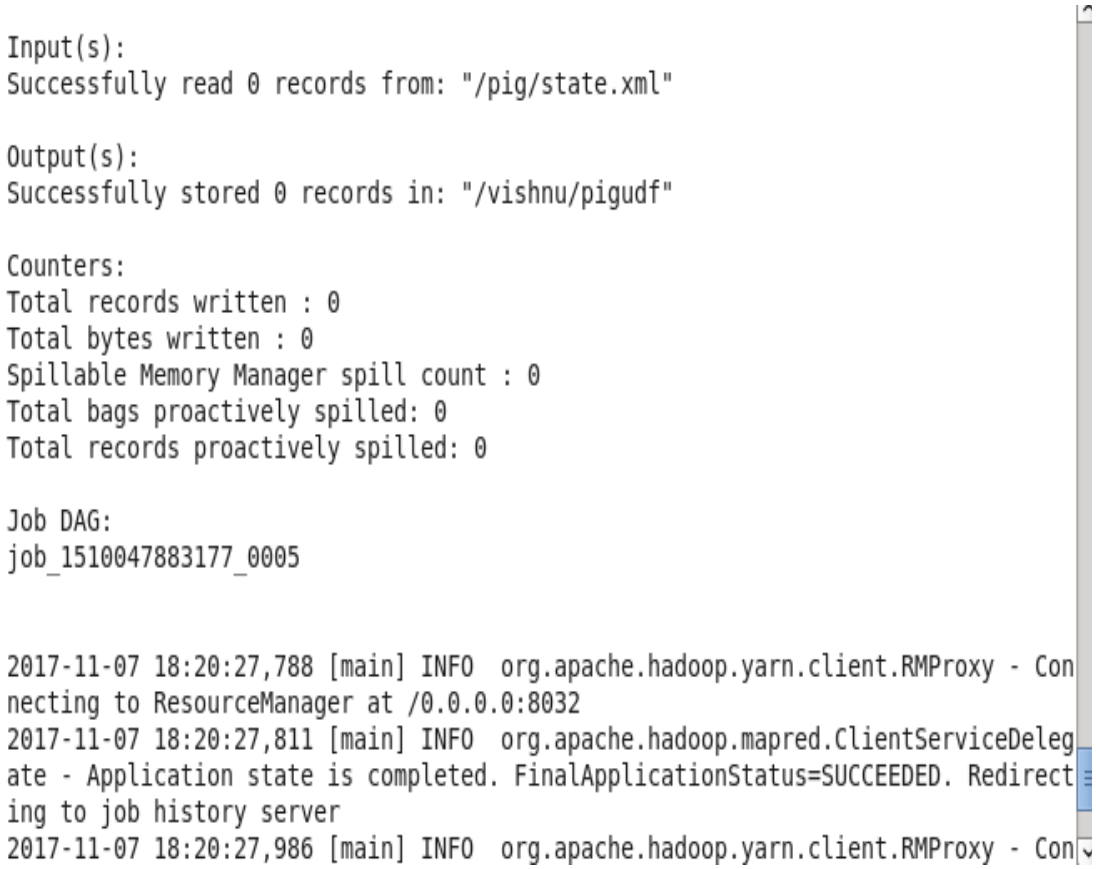
sqoop export --connect jdbc:mysql://localhost/state --username 'root' -P --table state80 --export-dir '/vishnu/pigudf/part-m-00000' --input-fields-terminated-by ',' -m 1

mysql> select * from state80;

State	district	BPL	total
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 |
+-------------------+---------------------------------------+-------+-------+
24 rows in set (0.00 sec)
```

Scree shots



```
[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                                      ,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
[acadgild@localhost ~]$
```

```
mysql> 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 |
+-------------------+-----------------------------------------+-------+-------+
24 rows in set (0.00 sec)
```