**State-Wise Development Analysis In India**

**Problem Statement**

**Problem statement 1:**

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

Export the results to mysql using sqoop

**Problem statement 2:**

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

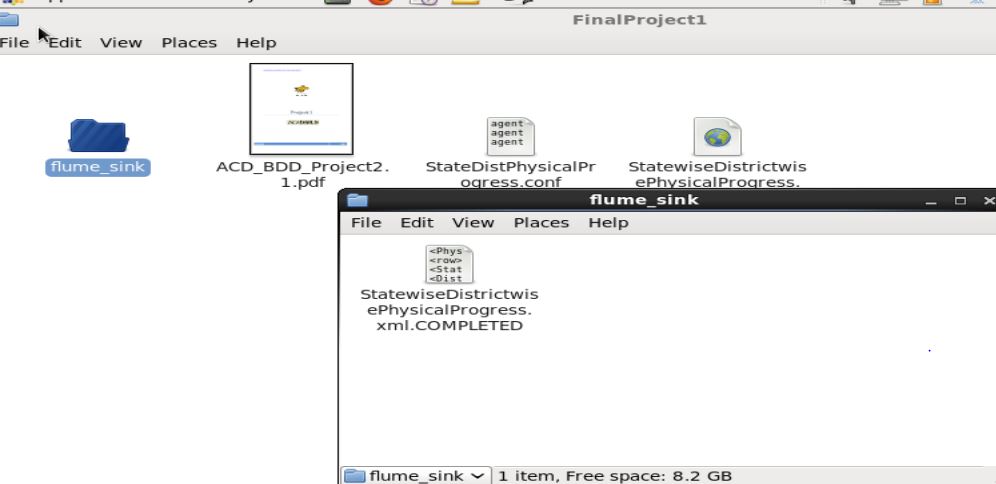
Export the results to mysql using sqoop.

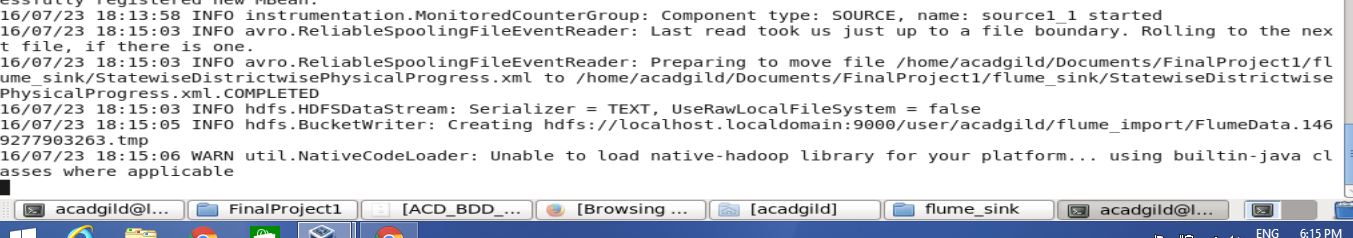
**Solution**

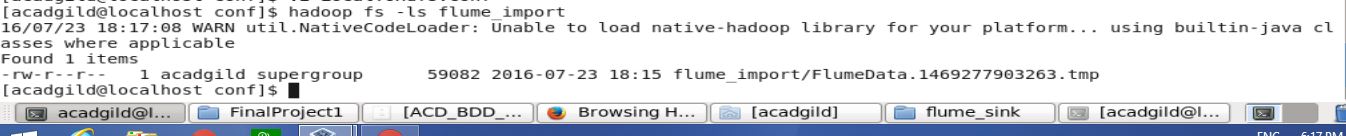
flume-ng agent -n agent1 -f /usr/local/flume/conf/LocalToHdfs.conf

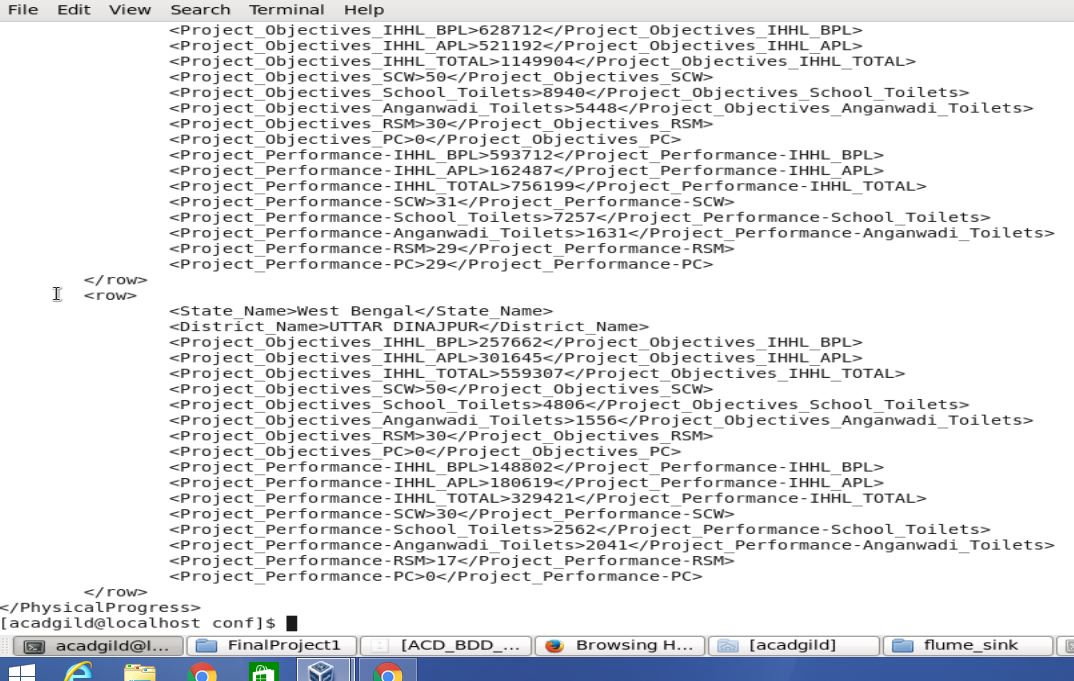












**Problem statement 1:**

register /home/acadgild/Documents/FinalProject1/pig\_xml.jar

xml\_input\_data = LOAD 'flume\_import/FlumeData.1469277903263' using pig.XML.newloader('row') as (R:chararray);

parsed\_data = FOREACH xml\_input\_data GENERATE FLATTEN(REGEX\_EXTRACT\_ALL(R,

'<row>\\s\*<State\_Name>(.\*)</State\_Name>\\s\*<District\_Name>(.\*)</District\_Name>\\s\*<Project\_Objectives\_IHHL\_BPL>(.\*)</Project\_Objectives\_IHHL\_BPL>\\s\*<Project\_Objectives\_IHHL\_APL>(.\*)</Project\_Objectives\_IHHL\_APL>\\s\*<Project\_Objectives\_IHHL\_TOTAL>(.\*)</Project\_Objectives\_IHHL\_TOTAL>\\s\*<Project\_Objectives\_SCW>(.\*)</Project\_Objectives\_SCW>\\s\*<Project\_Objectives\_School\_Toilets>(.\*)</Project\_Objectives\_School\_Toilets>\\s\*<Project\_Objectives\_Anganwadi\_Toilets>(.\*)</Project\_Objectives\_Anganwadi\_Toilets>\\s\*<Project\_Objectives\_RSM>(.\*)</Project\_Objectives\_RSM>\\s\*<Project\_Objectives\_PC>(.\*)</Project\_Objectives\_PC>\\s\*<Project\_Performance-IHHL\_BPL>(.\*)</Project\_Performance-IHHL\_BPL>\\s\*<Project\_Performance-IHHL\_APL>(.\*)</Project\_Performance-IHHL\_APL>\\s\*<Project\_Performance-IHHL\_TOTAL>(.\*)</Project\_Performance-IHHL\_TOTAL>\\s\*<Project\_Performance-SCW>(.\*)</Project\_Performance-SCW>\\s\*<Project\_Performance-School\_Toilets>(.\*)</Project\_Performance-School\_Toilets>\\s\*<Project\_Performance-Anganwadi\_Toilets>(.\*)</Project\_Performance-Anganwadi\_Toilets>\\s\*<Project\_Performance-RSM>(.\*)</Project\_Performance-RSM>\\s\*<Project\_Performance-PC>(.\*)</Project\_Performance-PC>\\s\*</row>'));

DUMP parsed\_data;

STORE parsed\_data INTO 'parsed\_data\_dump';

input\_data = LOAD 'parsed\_data\_dump/part-m-00000' USING PigStorage('\t') AS (f1:chararray, f2:chararray, f3:int, f4:int, f5:int, f6:int, f7:int, f8:int, f9:int, f10:int, f11:int, f12:int, f13:int, f14:int, f15:int, f16:int, f17:int, f18:int);

extract\_col = FOREACH input\_data GENERATE f2,f3,f11;

filter\_col = FILTER extract\_col BY f3==f11;

DUMP filter\_col;

STORE filter\_col INTO 'final\_data\_dump' USING PigStorage(',');

Problem 1 output:

NIZAMABAD,225519,225519

TIRAP,5780,5780

HAILAKANDI,49837,49837

MADHUBANI,67482,67482

NORTH GOA,15000,15000

AHMEDABAD,80192,80192

DANGS,27900,27900

NAVSARI,75015,75015

PORBANDAR,17024,17024

SURAT,158797,158797

FARIDABAD,22254,22254

HISAR,46463,46463

JHAJJAR,22014,22014

MAHENDRAGARH,17500,17500

PANCHKULA,8760,8760

PANIPAT,28000,28000

ROHTAK,22171,22171

SIRSA,35400,35400

HAMIRPUR,11593,11593

KINNAUR,1560,1560

KULLU,9989,9989

LAHAUL &amp; SPITI,2413,2413

SHIMLA,23874,23874

SOLAN,10858,10858

UNA,8360,8360

DEOGHAR,75153,75153

LOHARDAGA,22626,22626

HASSAN,64134,64134

MANGALORE(DAKSHINA KANNADA),59478,59478

UDUPI,52348,52348

ALAPPUZHA,114359,114359

KOLLAM,95130,95130

KOTTAYAM,28118,28118

KOZHIKODE,42285,42285

PALAKKAD,107018,107018

PATHANAMTHITTA,53799,53799

WAYANAD,50655,50655

GADCHIROLI,75900,75900

SINDHUDURG,43874,43874

WEST GARO HILLS,44385,44385

CHAMPHAI,11077,11077

LAWNGTLAI,16544,16544

HANUMANGARH,31621,31621

ERODE,165306,165306

KARUR,105280,105280

NAMAKKAL,117538,117538

TIRUCHIRAPPALLI,77747,77747

TIRUVANNAMALAI,209116,209116

DHALAI,53507,53507

SOUTH TRIPURA,139456,139456

WEST TRIPURA,183405,183405

AMBEDKAR NAGAR,132725,132725

BALRAMPUR,65273,65273

BAREILLY,110000,110000

BIJNOR,110403,110403

BUDAUN,107603,107603

ETAWAH,94097,94097

FARRUKHABAD,120471,120471

FIROZABAD,19843,19843

GHAZIABAD,10810,10810

HARDOI,199989,199989

JYOTIBA PHULE NAGAR,48008,48008

LUCKNOW,113188,113188

MAHARAJGANJ,145090,145090

MAHOBA,53117,53117

MORADABAD,76018,76018

MUZAFFARNAGAR,51660,51660

PILIBHIT,95178,95178

SONBHADRA,138370,138370

SULTANPUR,168843,168843

**Sqoop job to import data from hdfs to mysql table:**

sqoop export --connect jdbc:mysql://localhost/mysample --table dist\_bpl\_full -m 1 --export-dir final\_data\_dump/part-m-00000

mysql> select \* from dist\_bpl\_full;

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

| dist\_name | obj\_bpl | perf\_bpl |

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

| NIZAMABAD | 225519 | 225519 |

| TIRAP | 5780 | 5780 |

| HAILAKANDI | 49837 | 49837 |

| MADHUBANI | 67482 | 67482 |

| NORTH GOA | 15000 | 15000 |

| AHMEDABAD | 80192 | 80192 |

| DANGS | 27900 | 27900 |

| NAVSARI | 75015 | 75015 |

| PORBANDAR | 17024 | 17024 |

| SURAT | 158797 | 158797 |

| FARIDABAD | 22254 | 22254 |

| HISAR | 46463 | 46463 |

| JHAJJAR | 22014 | 22014 |

| MAHENDRAGARH | 17500 | 17500 |

| PANCHKULA | 8760 | 8760 |

| PANIPAT | 28000 | 28000 |

| ROHTAK | 22171 | 22171 |

| SIRSA | 35400 | 35400 |

| HAMIRPUR | 11593 | 11593 |

| KINNAUR | 1560 | 1560 |

| KULLU | 9989 | 9989 |

| LAHAUL &amp; SPITI | 2413 | 2413 |

| SHIMLA | 23874 | 23874 |

| SOLAN | 10858 | 10858 |

| UNA | 8360 | 8360 |

| DEOGHAR | 75153 | 75153 |

| LOHARDAGA | 22626 | 22626 |

| HASSAN | 64134 | 64134 |

| MANGALORE(DAKSHINA KANNADA) | 59478 | 59478 |

| UDUPI | 52348 | 52348 |

| ALAPPUZHA | 114359 | 114359 |

| KOLLAM | 95130 | 95130 |

| KOTTAYAM | 28118 | 28118 |

| KOZHIKODE | 42285 | 42285 |

| PALAKKAD | 107018 | 107018 |

| PATHANAMTHITTA | 53799 | 53799 |

| WAYANAD | 50655 | 50655 |

| GADCHIROLI | 75900 | 75900 |

| SINDHUDURG | 43874 | 43874 |

| WEST GARO HILLS | 44385 | 44385 |

| CHAMPHAI | 11077 | 11077 |

| LAWNGTLAI | 16544 | 16544 |

| HANUMANGARH | 31621 | 31621 |

| ERODE | 165306 | 165306 |

| KARUR | 105280 | 105280 |

| NAMAKKAL | 117538 | 117538 |

| TIRUCHIRAPPALLI | 77747 | 77747 |

| TIRUVANNAMALAI | 209116 | 209116 |

| DHALAI | 53507 | 53507 |

| SOUTH TRIPURA | 139456 | 139456 |

| WEST TRIPURA | 183405 | 183405 |

| AMBEDKAR NAGAR | 132725 | 132725 |

| BALRAMPUR | 65273 | 65273 |

| BAREILLY | 110000 | 110000 |

| BIJNOR | 110403 | 110403 |

| BUDAUN | 107603 | 107603 |

| ETAWAH | 94097 | 94097 |

| FARRUKHABAD | 120471 | 120471 |

| FIROZABAD | 19843 | 19843 |

| GHAZIABAD | 10810 | 10810 |

| HARDOI | 199989 | 199989 |

| JYOTIBA PHULE NAGAR | 48008 | 48008 |

| LUCKNOW | 113188 | 113188 |

| MAHARAJGANJ | 145090 | 145090 |

| MAHOBA | 53117 | 53117 |

| MORADABAD | 76018 | 76018 |

| MUZAFFARNAGAR | 51660 | 51660 |

| PILIBHIT | 95178 | 95178 |

| SONBHADRA | 138370 | 138370 |

| SULTANPUR | 168843 | 168843 |

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

70 rows in set (0.00 sec)

**Problem statement 2:**

Pig UDF:

package percent;

import java.io.IOException;

import org.apache.pig.EvalFunc;

import org.apache.pig.data.Tuple;

public class Percentage extends EvalFunc<Float>{

public Float exec(Tuple input) throws IOException {

if(input == null || input.size() == 0){

return null;

}

// TODO Auto-generated method stub

float first = (Integer)input.get(0);

float second = (Integer)input.get(1);

try{

float diff = (first - second);

float p = diff/first;

float prct = p\*100;

if(prct<80){

return null;

}

return prct;

} catch (Exception e){

throw new IOException("Crash!", e);

}

}

}

Pig Script:

REGISTER /home/acadgild/workspace/percent.jar

input\_data = LOAD 'parsed\_data\_dump/part-m-00000' USING PigStorage('\t') AS (f1:chararray, f2:chararray, f3:int, f4:int, f5:int, f6:int, f7:int, f8:int, f9:int, f10:int, f11:int, f12:int, f13:int, f14:int, f15:int, f16:int, f17:int, f18:int);

extract\_col = FOREACH input\_data GENERATE f2,f3,f11;

B = FOREACH extract\_col GENERATE f2,percent.Percentage(f3,f11);

C = FILTER B BY NOT($1 is NULL);

DUMP C;

STORE C INTO 'Proj1\_sol2' USING PigStorage(',');

Output of STORE:

BONGAIGAON,89.89967

KARBI ANGLONG,84.12219

SITAMARHI,84.3256

DADRA AND NAGAR HAVELI,98.508064

RAMBAN ,87.6868

REASI ,84.47442

SAMBA ,89.11565

UKHRUL,89.90818

KIPHIRE,100.0

LONGLENG,100.0

PONDICHERRY,87.4

AMRITSAR,96.99519

MOGA,93.91714

PATIALA,83.099495

RUPNAGAR,85.10503

Tarn Taran,80.80855

BHARATPUR,83.923164

**Sqoop job:**

sqoop export --connect jdbc:mysql://localhost/mysample --table dist\_bpl\_eighty -m 1 --export-dir Proj1\_sol2/part-m-00000

mysql> select \* from dist\_bpl\_eighty;

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

| dist\_name | ach\_bpl |

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

| BONGAIGAON | 89.8997 |

| KARBI ANGLONG | 84.1222 |

| SITAMARHI | 84.3256 |

| DADRA AND NAGAR HAVELI | 98.5081 |

| RAMBAN | 87.6868 |

| REASI | 84.4744 |

| SAMBA | 89.1156 |

| UKHRUL | 89.9082 |

| KIPHIRE | 100 |

| LONGLENG | 100 |

| PONDICHERRY | 87.4 |

| AMRITSAR | 96.9952 |

| MOGA | 93.9171 |

| PATIALA | 83.0995 |

| RUPNAGAR | 85.105 |

| Tarn Taran | 80.8085 |

| BHARATPUR | 83.9232 |

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

17 rows in set (0.01 sec)

