

Acadgild – Hadoop and Spark Mini Project Assignment

The problem statement:

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

Export the results to mysql using sqoop

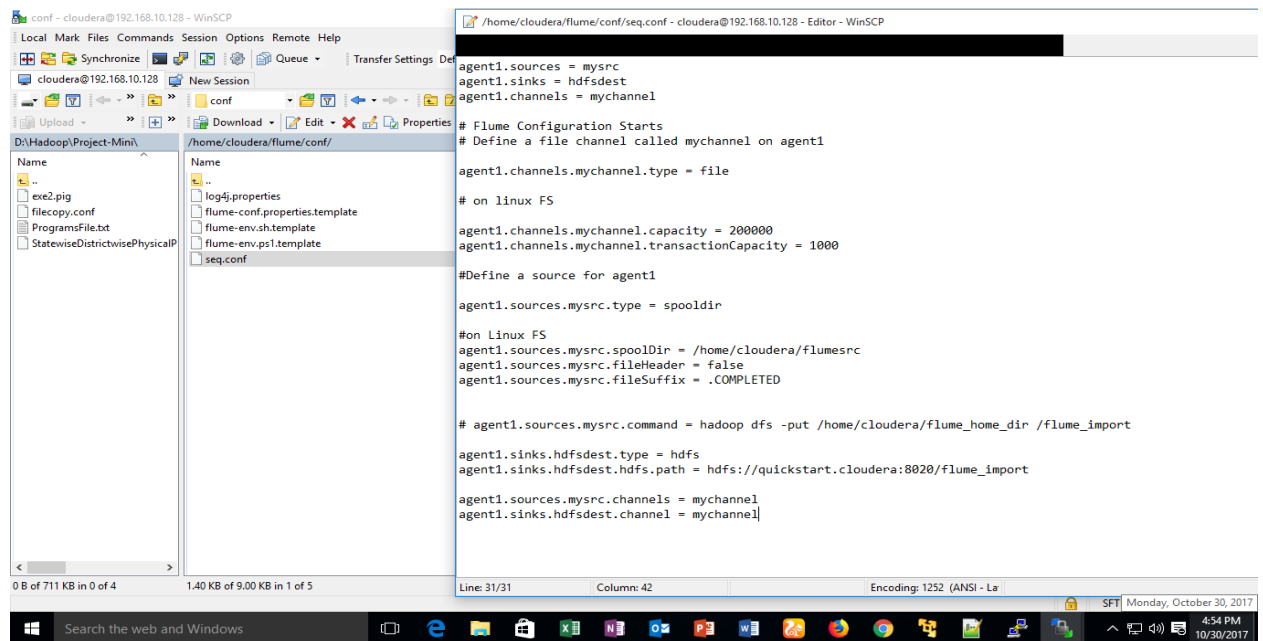
2. Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards.

Export the results to MySQL using Sqoop.

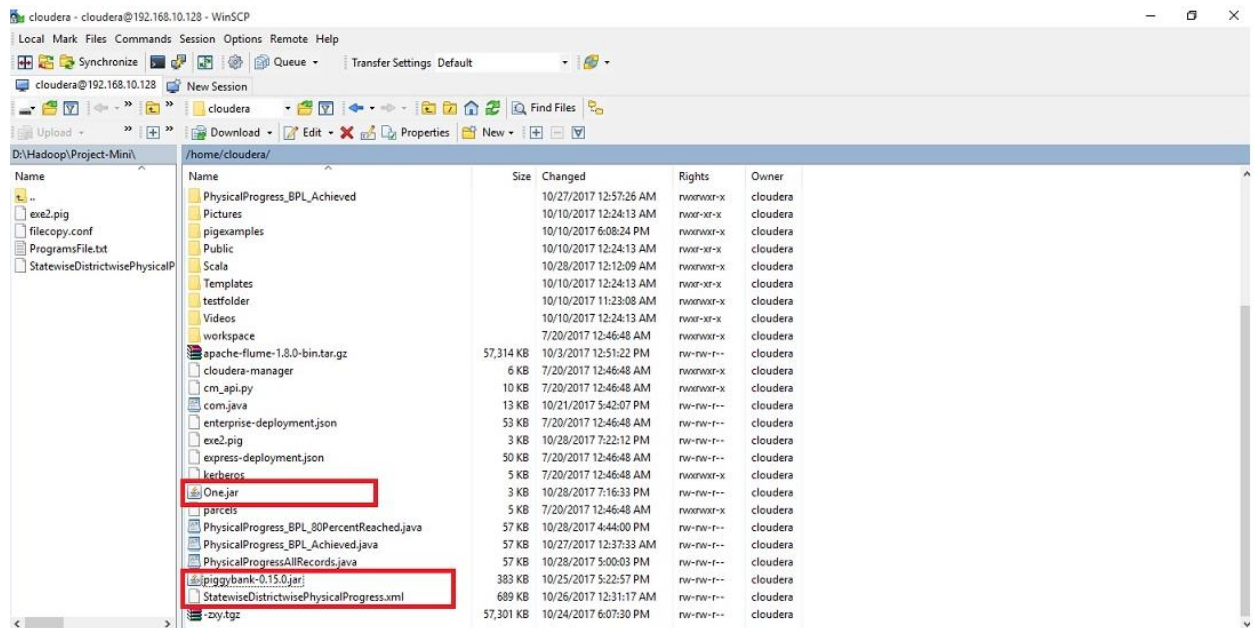
2. Screenshot of the command executed:

The solution of the above problem statements contains different commands and java, those are:

- Flume conf configuration file screen shot (Which is common to Project Statement 1 & 2)



- WinSCP screen shot to depict the files which are moved to the linux directory (Which is common to Project Statement 1 & 2)

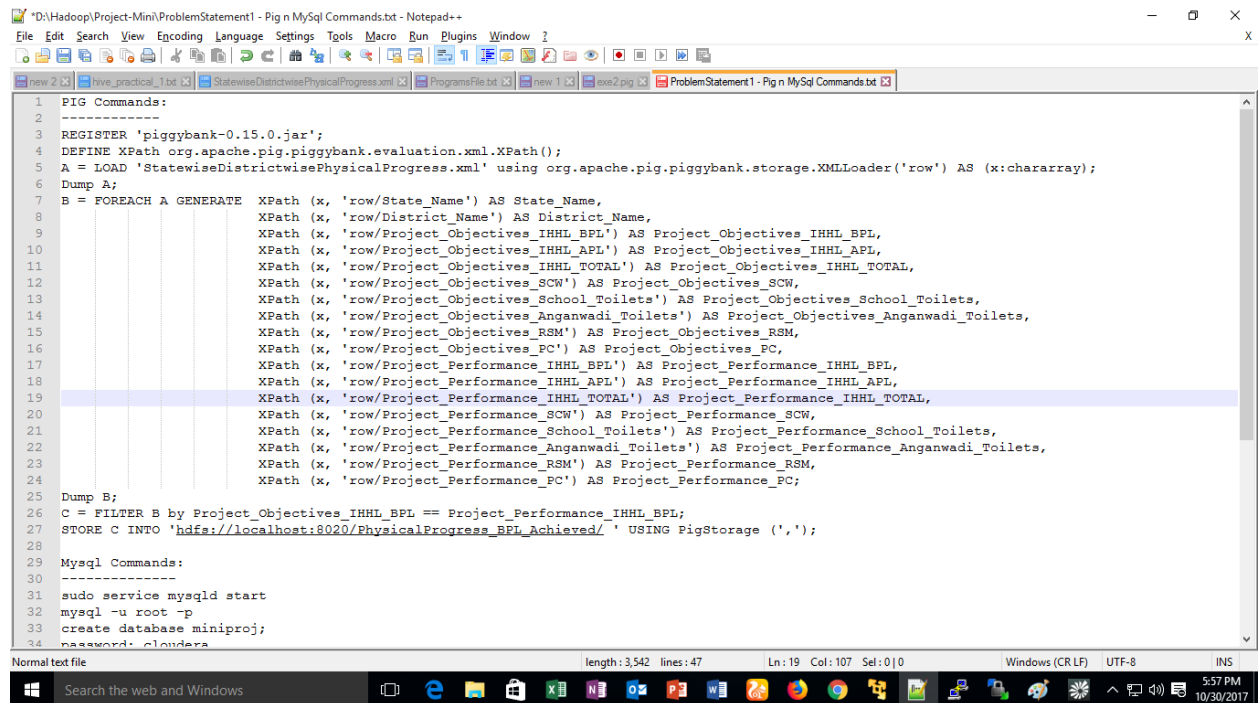


One.jar – contains the UDF implementation for the Problem statements – Point 2.

Piggybank-0.15.0.jar – which contains the classes for the xml file parsing.

StatewiseDistrictwisePhysicalProgress.xml – which contains the sample data

c. Project Statement 1 – Pig and MySQL Commands Screen shot:



```
"D:\Hadoop\Project-Mini\ProblemStatement1 - Pig n MySQL Commands.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

new 2 x hve_practical_1.txt StatewiseDistrictwisePhysicalProgress.xml ProgramsFile.txt new 1 x exe2.pig ProblemStatement1 - Pig n MySQL Commands.txt

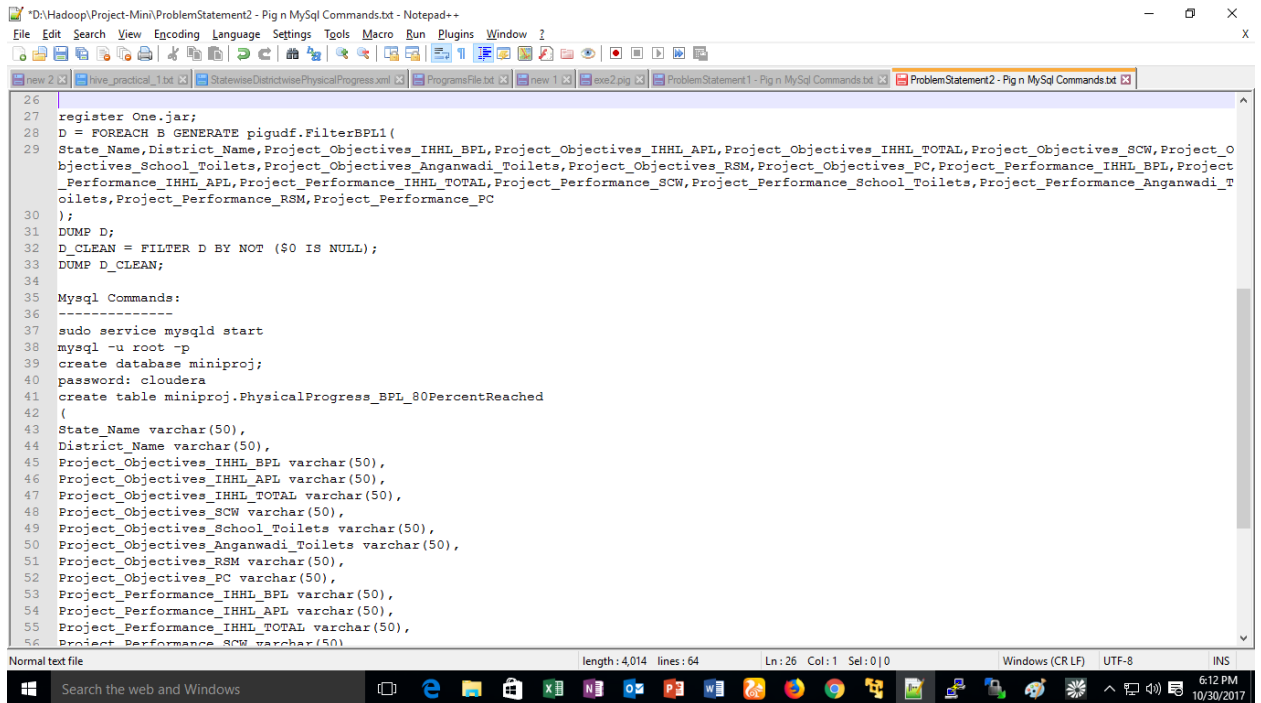
20 XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,
21 XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_School_Toilets,
22 XPath (x, 'row/Project_Performance_Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,
23 XPath (x, 'row/Project_Performance_RSM') AS Project_Performance_RSM,
24 XPath (x, 'row/Project_Performance_PC') AS Project_Performance_PC;
25
26 Dump B;
27 C = FILTER B by Project_Objectives_IHHL_BPL == Project_Performance_IHHL_BPL;
28 STORE C INTO 'hdfs://localhost:8020/PhysicalProgress_BPL_Achieved/' USING PigStorage(',');
29
30 Mysql Commands:
31 -----
32 sudo service mysqld start
33 mysql -u root -p
34 create database miniproj;
35 password: cloudera
36
37 create table miniproj.PhysicalProgress_BPL_Achieved
38 (
39 State_Name varchar(50), District_Name varchar(50), Project_Objectives_IHHL_BPL varchar(50), Project_Objectives_IHHL_APL varchar(50),
40 Project_Objectives_IHHL_TOTAL varchar(50), Project_Objectives_SCW varchar(50), Project_Objectives_School_Toilets varchar(50),
41 Project_Objectives_Anganwadi_Toilets varchar(50), Project_Objectives_RSM varchar(50), Project_Objectives_PC varchar(50),
42 Project_Performance_IHHL_BPL varchar(50), Project_Performance_IHHL_APL varchar(50), Project_Performance_IHHL_TOTAL varchar(50),
43 Project_Performance_SCW varchar(50), Project_Performance_School_Toilets varchar(50), Project_Performance_Anganwadi_Toilets varchar(50),
44 Project_Performance_RSM varchar(50), Project_Performance_PC varchar(50)
45 );
46
47 sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_Achieved' --export-dir
48 '/PhysicalProgress_BPL_Achieved' --input-fields-terminated-by ',' --m 1 --columns
49 State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O
bjectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project
_Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_T
oilets,Project_Performance_RSM,Project_Performance_PC;
49
Normal text file length: 3,544 lines: 48 Ln: 45 Col: 1 Sel: 0 | 0 Windows (CR LF) UTF-8 INS
```

d. Project Statement 2 – Pig and MySQL Commands Screen shot:

```
"D:\Hadoop\Project-Mini\ProblemStatement2 - Pig n MySQL Commands.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

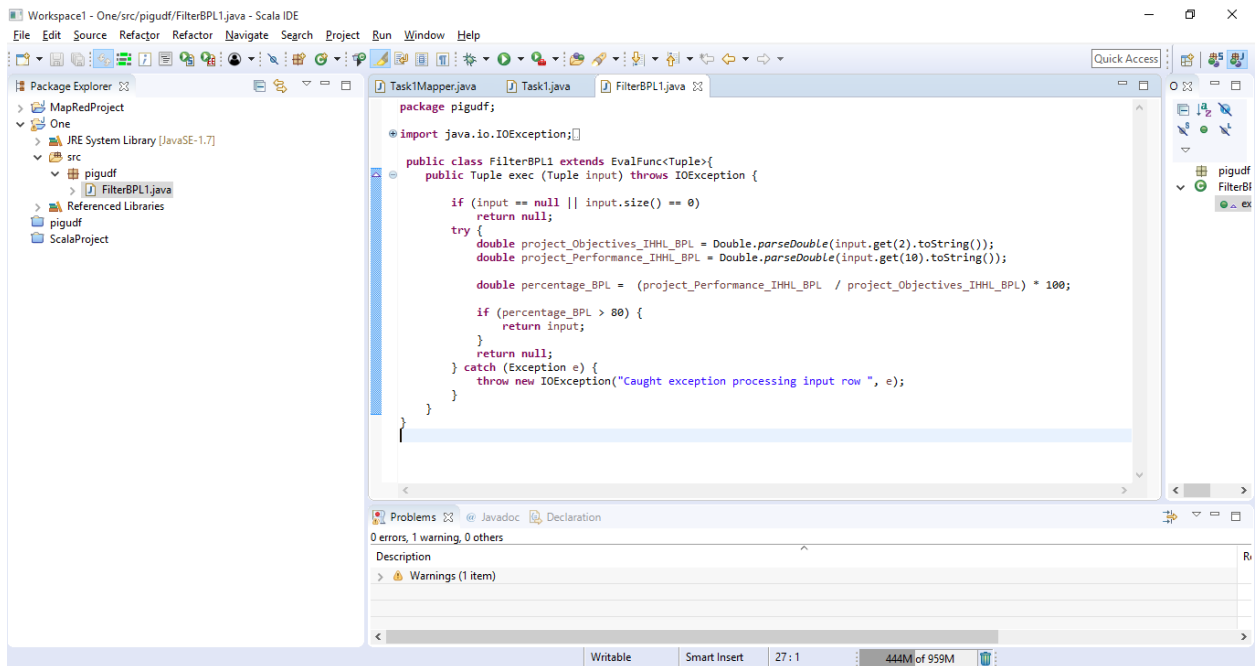
new 2 x hve_practical_1.txt StatewiseDistrictwisePhysicalProgress.xml ProgramsFile.txt new 1 x exe2.pig ProblemStatement1 - Pig n MySQL Commands.txt ProblemStatement2 - Pig n MySQL Commands.txt

1 PIG Commands:
2 -----
3 REGISTER 'piggybank-0.15.0.jar';
4 DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath();
5 A = LOAD 'StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') AS (x:chararray);
6
7 Dump A;
8
9 B = FOREACH A GENERATE XPath (x, 'row/State_Name') AS State_Name,
10 XPath (x, 'row/District_Name') AS District_Name,
11 XPath (x, 'row/Project_Objectives_IHHL_BPL') AS Project_Objectives_IHHL_BPL,
12 XPath (x, 'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,
13 XPath (x, 'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_TOTAL,
14 XPath (x, 'row/Project_Objectives_SCW') AS Project_Objectives_SCW,
15 XPath (x, 'row/Project_Objectives_School_Toilets') AS Project_Objectives_School_Toilets,
16 XPath (x, 'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_Anganwadi_Toilets,
17 XPath (x, 'row/Project_Objectives_RSM') AS Project_Objectives_RSM,
18 XPath (x, 'row/Project_Objectives_PC') AS Project_Objectives_PC,
19 XPath (x, 'row/Project_Performance_IHHL_BPL') AS Project_Performance_IHHL_BPL,
20 XPath (x, 'row/Project_Performance_IHHL_APL') AS Project_Performance_IHHL_APL,
21 XPath (x, 'row/Project_Performance_IHHL_TOTAL') AS Project_Performance_IHHL_TOTAL,
22 XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,
23 XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_School_Toilets,
24 XPath (x, 'row/Project_Performance_Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,
25 XPath (x, 'row/Project_Performance_RSM') AS Project_Performance_RSM,
26 XPath (x, 'row/Project_Performance_PC') AS Project_Performance_PC;
27
28 Dump B;
29
30 register One.jar;
31 D = FOREACH B GENERATE pigudf.FilterBPL1(
32 State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O
bjectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project
_Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_T
oilets,Project_Performance_RSM,Project_Performance_PC
33 );
34
35 DUMP D;
36
Normal text file length: 4,018 lines: 66 Ln: 29 Col: 430 Sel: 0 | 0 Windows (CR LF) UTF-8 INS
Search the web and Windows 6:11 PM 10/30/2017
```



```
26 register One.jar;
27 D = FOREACH B GENERATE pigudf.FilterBPL1(
28 State_Name, District_Name, Project_Objectives_IHHL_BPL, Project_Objectives_IHHL_APL, Project_Objectives_IHHL_TOTAL, Project_Objectives_SCW, Project_O
29 bjectives_School_Toilets, Project_Objectives_Anganwadi_Toilets, Project_Objectives_RSM, Project_Objectives_PC, Project_Performance_IHHL_BPL, Project
_Performance_IHHL_APL, Project_Performance_IHHL_TOTAL, Project_Performance_SCW, Project_Performance_School_Toilets, Project_Performance_Anganwadi_T
oilets, Project_Performance_RSM, Project_Performance_PC
30 );
31 DUMP D;
32 D_CLEAN = FILTER D BY NOT ($0 IS NULL);
33 DUMP D_CLEAN;
34
35 Mysql Commands:
36 -----
37 sudo service mysqld start
38 mysql -u root -p
39 create database miniproj;
40 password: cloudera
41 create table miniproj.PhysicalProgress_BPL_80PercentReached
42 (
43 State_Name varchar(50),
44 District_Name varchar(50),
45 Project_Objectives_IHHL_BPL varchar(50),
46 Project_Objectives_IHHL_APL varchar(50),
47 Project_Objectives_IHHL_TOTAL varchar(50),
48 Project_Objectives_SCW varchar(50),
49 Project_Objectives_School_Toilets varchar(50),
50 Project_Objectives_Anganwadi_Toilets varchar(50),
51 Project_Objectives_RSM varchar(50),
52 Project_Objectives_PC varchar(50),
53 Project_Performance_IHHL_BPL varchar(50),
54 Project_Performance_IHHL_APL varchar(50),
55 Project_Performance_IHHL_TOTAL varchar(50),
56 Project_Performance_SCW varchar(50),
57 Project_Performance_School_Toilets varchar(50),
58 Project_Performance_Anganwadi_Toilets varchar(50),
59 Project_Performance_RSM varchar(50),
60 Project_Performance_PC varchar(50)
61 );
62 sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir
'/PhysicalProgress_BPL_80PercentReached' --input-fields-terminated-by ',' -m 1 --columns
State_Name, District_Name, Project_Objectives_IHHL_BPL, Project_Objectives_IHHL_APL, Project_Objectives_IHHL_TOTAL, Project_Objectives_SCW, Project_O
bjectives_School_Toilets, Project_Objectives_Anganwadi_Toilets, Project_Objectives_RSM, Project_Objectives_PC, Project_Performance_IHHL_BPL, Project
_Performance_IHHL_APL, Project_Performance_IHHL_TOTAL, Project_Performance_SCW, Project_Performance_School_Toilets, Project_Performance_Anganwadi_T
oilets, Project_Performance_RSM, Project_Performance_PC;
```

- e. Project Statement 2 – Screen shot of the UDF Function Screen shot which is implemented in Java under the project named “One” with the class name as “FilterBPL1.java:



- f. Mysql command prompt screen – to execute the Mysql related command
- g. Linux command prompt screen – to execute the Linux commands and Hadoop commands
- h. Pig (Grunt shell) command prompt screens – to execute the pig commands.

3. Explanation of the code and its working:

a.

Code Explanation and Way of Working – Common to Problem statement 1 & 2.

- i) Set up the Flume directory where flume job will be triggered whenever any new file comes to “/home/cloudera/flumesrc” linux directory
- ii) Once the new file copied into the “hdfs://quickstart.cloudera:8020/flume_import” sink directory, then in the source folder it will rename the copied file with the extension of .COMPLETED.
- iii) For this flume configuration file’s content, please refer the uploaded file seq.conf file in the GITHUB.
- iv) flume-ng agent -n agent1 -f conf/seq.conf – This command will be start the flume job

b.

Code Explanation and Way of Working – Common to Problem statement 1. Below screen shot explains the flow and purpose those commands:

```

PIG Commands:
-----
REGISTER 'piggybank-0.15.0.jar';
DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath();
A = LOAD 'StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') AS (x:chararray);
Dump A;
B = FOREACH A GENERATE XPath (x, 'row/State_Name') AS State_Name,
XPath (x, 'row/District_Name') AS District_Name,
XPath (x, 'row/Project_Objectives_IHHL_BPL') AS Project_Objectives_IHHL_BPL,
XPath (x, 'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,
XPath (x, 'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_TOTAL,
XPath (x, 'row/Project_Objectives_SCW') AS Project_Objectives_SCW,
XPath (x, 'row/Project_Objectives_School_Toilets') AS Project_Objectives_School_Toilets,
XPath (x, 'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_Anganwadi_Toilets,
XPath (x, 'row/Project_Objectives_RSM') AS Project_Objectives_RSM,
XPath (x, 'row/Project_Objectives_PC') AS Project_Objectives_PC,
XPath (x, 'row/Project_Performance_IHHL_BPL') AS Project_Performance_IHHL_BPL,
XPath (x, 'row/Project_Performance_IHHL_APL') AS Project_Performance_IHHL_APL,
XPath (x, 'row/Project_Performance_IHHL_TOTAL') AS Project_Performance_IHHL_TOTAL,
XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,
XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_School_Toilets,
XPath (x, 'row/Project_Performance_Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,
XPath (x, 'row/Project_Performance_RSM') AS Project_Performance_RSM,
XPath (x, 'row/Project_Performance_PC') AS Project_Performance_PC;
Dump B;
C = FILTER B by Project_Objectives_IHHL_BPL == Project_Performance_IHHL_BPL;
STORE C INTO 'hdfs://localhost:8020/PhysicalProgress_BPL_Achieved/' USING PigStorage ('');

```

```

Mysql Commands:
-----
sudo service mysqld start
mysql -u root -p
create database miniproj;
password: cloudera

create table miniproj.PhysicalProgress_BPL_Achieved
(
State_Name varchar(50), District_Name varchar(50), Project_Objectives_IHHL_BPL varchar(50), Project_Objectives_IHHL_APL varchar(50),
Project_Objectives_IHHL_TOTAL varchar(50), Project_Objectives_SCW varchar(50), Project_Objectives_School_Toilets varchar(50),
Project_Objectives_Anganwadi_Toilets varchar(50), Project_Objectives_RSM varchar(50), Project_Objectives_PC varchar(50),
Project_Performance_IHHL_BPL varchar(50), Project_Performance_IHHL_APL varchar(50), Project_Performance_IHHL_TOTAL varchar(50),
Project_Performance_SCW varchar(50), Project_Performance_School_Toilets varchar(50), Project_Performance_Anganwadi_Toilets varchar(50),
Project_Performance_RSM varchar(50), Project_Performance_PC varchar(50)
);

```

C.

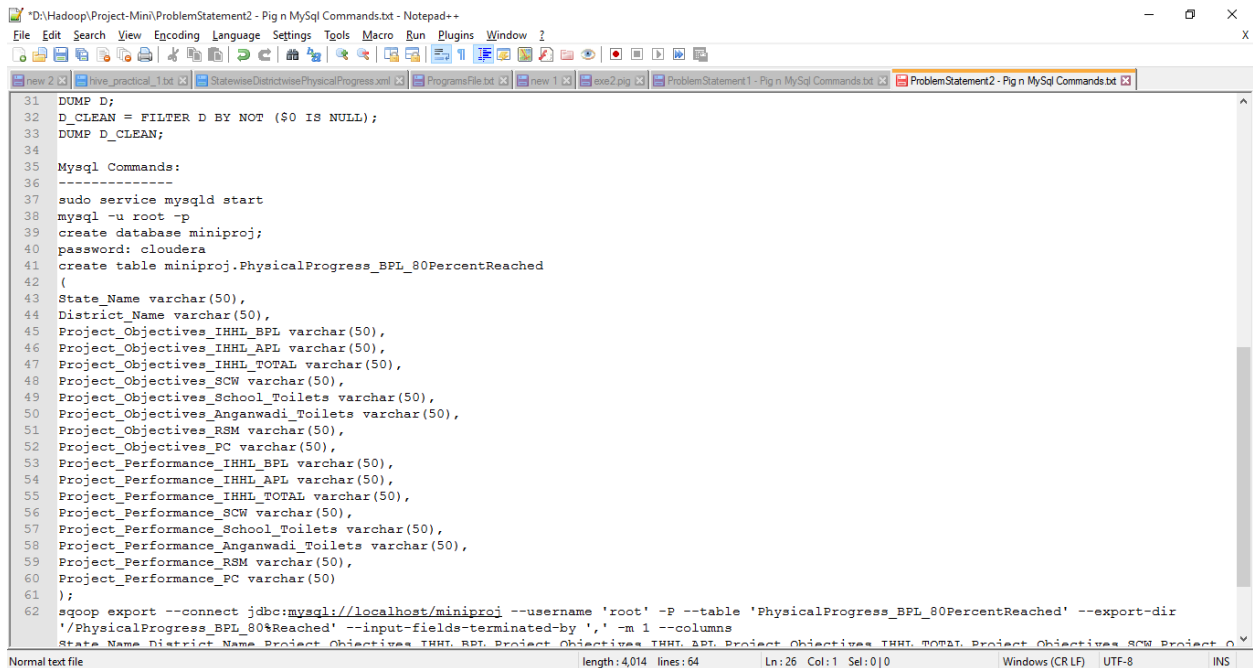
Code Explanation and Way of Working – Common to Problem statement2. Below screen shot explains the flow and purpose those commands:

```

1 PIG Commands:
2 -----
3 REGISTER 'piggybank-0.15.0.jar';
4 DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath();
5 A = LOAD 'StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') AS (x:chararray);
6 Dump A;
7 B = FOREACH A GENERATE XPath (x, 'row/State_Name') AS State_Name,
8 XPath (x, 'row/District_Name') AS District_Name,
9 XPath (x, 'row/Project_Objectives_IHHL_BPL') AS Project_Objectives_IHHL_BPL,
10 XPath (x, 'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,
11 XPath (x, 'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_TOTAL,
12 XPath (x, 'row/Project_Objectives_SCW') AS Project_Objectives_SCW,
13 XPath (x, 'row/Project_Objectives_School_Toilets') AS Project_Objectives_School_Toilets,
14 XPath (x, 'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_Anganwadi_Toilets,
15 XPath (x, 'row/Project_Objectives_RSM') AS Project_Objectives_RSM,
16 XPath (x, 'row/Project_Objectives_PC') AS Project_Objectives_PC,
17 XPath (x, 'row/Project_Performance_IHHL_BPL') AS Project_Performance_IHHL_BPL,
18 XPath (x, 'row/Project_Performance_IHHL_APL') AS Project_Performance_IHHL_APL,
19 XPath (x, 'row/Project_Performance_IHHL_TOTAL') AS Project_Performance_IHHL_TOTAL,
20 XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,
21 XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_School_Toilets,
22 XPath (x, 'row/Project_Performance_Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,
23 XPath (x, 'row/Project_Performance_RSM') AS Project_Performance_RSM,
24 XPath (x, 'row/Project_Performance_PC') AS Project_Performance_PC;
25 Dump B;
26
27 register One.jar;
28 D = FOREACH B GENERATE pigudf.FilterBPL(
29 State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O
30 bjectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project
31 _Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_T
32 oilets,Project_Performance_RSM,Project_Performance_PC
33 );
34 DUMP D;

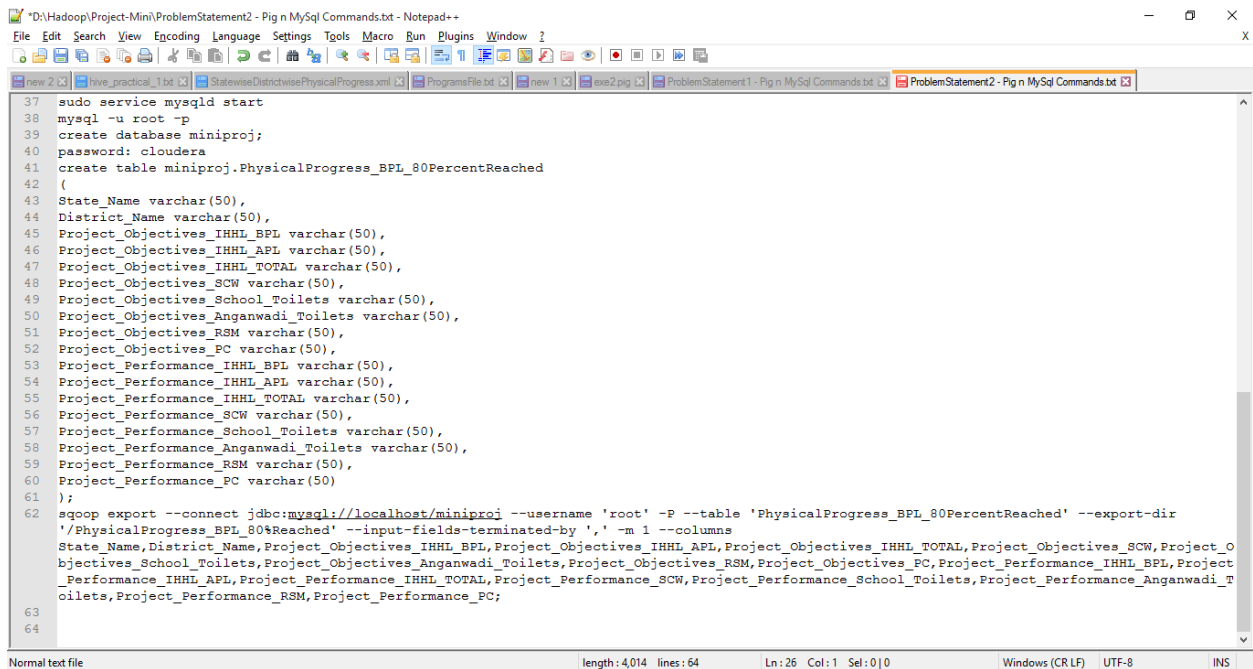
```

Normal text file length: 4,014 lines: 64 Ln: 26 Col: 1 Sel: 0 | 0 Windows (CRLF) UTF-8 INS



```
31 DUMP D;  
32 D_CLEAN = FILTER D BY NOT ($0 IS NULL);  
33 DUMP D_CLEAN;  
34  
35 Mysql Commands:  
36 -----  
37 sudo service mysqld start  
38 mysql -u root -p  
39 create database miniproj;  
40 password: cloudera  
41 create table miniproj.PhysicalProgress_BPL_80PercentReached  
42 (  
43   State_Name varchar(50),  
44   District_Name varchar(50),  
45   Project_Objectives_IHHL_BPL varchar(50),  
46   Project_Objectives_IHHL_APL varchar(50),  
47   Project_Objectives_IHHL_TOTAL varchar(50),  
48   Project_Objectives_SCW varchar(50),  
49   Project_Objectives_School_Toilets varchar(50),  
50   Project_Objectives_Anganwadi_Toilets varchar(50),  
51   Project_Objectives_RSM varchar(50),  
52   Project_Objectives_PC varchar(50),  
53   Project_Performance_IHHL_BPL varchar(50),  
54   Project_Performance_IHHL_APL varchar(50),  
55   Project_Performance_IHHL_TOTAL varchar(50),  
56   Project_Performance_SCW varchar(50),  
57   Project_Performance_School_Toilets varchar(50),  
58   Project_Performance_Anganwadi_Toilets varchar(50),  
59   Project_Performance_RSM varchar(50),  
60   Project_Performance_PC varchar(50)  
61 );  
62 sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir  
63   '/PhysicalProgress_BPL_80PercentReached' --input-fields-terminated-by ',' -m 1 --columns  
64   State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O
```

Normal text file length: 4,014 lines: 64 Ln: 26 Col: 1 Sel: 0 | 0 Windows (CR LF) UTF-8 INS



```
37 sudo service mysqld start  
38 mysql -u root -p  
39 create database miniproj;  
40 password: cloudera  
41 create table miniproj.PhysicalProgress_BPL_80PercentReached  
42 (  
43   State_Name varchar(50),  
44   District_Name varchar(50),  
45   Project_Objectives_IHHL_BPL varchar(50),  
46   Project_Objectives_IHHL_APL varchar(50),  
47   Project_Objectives_IHHL_TOTAL varchar(50),  
48   Project_Objectives_SCW varchar(50),  
49   Project_Objectives_School_Toilets varchar(50),  
50   Project_Objectives_Anganwadi_Toilets varchar(50),  
51   Project_Objectives_RSM varchar(50),  
52   Project_Objectives_PC varchar(50),  
53   Project_Performance_IHHL_BPL varchar(50),  
54   Project_Performance_IHHL_APL varchar(50),  
55   Project_Performance_IHHL_TOTAL varchar(50),  
56   Project_Performance_SCW varchar(50),  
57   Project_Performance_School_Toilets varchar(50),  
58   Project_Performance_Anganwadi_Toilets varchar(50),  
59   Project_Performance_RSM varchar(50),  
60   Project_Performance_PC varchar(50)  
61 );  
62 sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir  
63   '/PhysicalProgress_BPL_80PercentReached' --input-fields-terminated-by ',' -m 1 --columns  
64   State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O  
bjectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project_Per  
formance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_T  
oilets,Project_Performance_RSM,Project_Performance_PC;  
63  
64
```

Normal text file length: 4,014 lines: 64 Ln: 26 Col: 1 Sel: 0 | 0 Windows (CR LF) UTF-8 INS

4. Screenshot of the output

a.

Output of the Flume Job:

```
cloudera@quickstart:~  
-rw-r--r-- 1 cloudera supergroup 846 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363145  
-rw-r--r-- 1 cloudera supergroup 967 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363146  
-rw-r--r-- 1 cloudera supergroup 834 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363147  
-rw-r--r-- 1 cloudera supergroup 963 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363148  
-rw-r--r-- 1 cloudera supergroup 831 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363149  
-rw-r--r-- 1 cloudera supergroup 959 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363150  
-rw-r--r-- 1 cloudera supergroup 846 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363151  
-rw-r--r-- 1 cloudera supergroup 966 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363152  
-rw-r--r-- 1 cloudera supergroup 841 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363153  
-rw-r--r-- 1 cloudera supergroup 965 2017-10-30 07:23 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363154  
-rw-r--r-- 1 cloudera supergroup 182 2017-10-30 07:24 hdfs://quickstart  
.cloudera:8020/flume_import/FlumeData.1509373363155  
[cloudera@quickstart ~]$ hadoop fs -ls hdfs://quickstart.cloudera:8020/flume_im  
port/
```

b.

Screen shots for the output of Problem statement 1:

```
cloudera@quickstart:~  
login as: cloudera  
cloudera@192.168.10.129's password:  
Last login: Mon Oct 30 06:59:36 2017 from 192.168.10.1  
[cloudera@quickstart ~]$ pig -x local
```



```
cloudera@quickstart:~  
(West Bengal,MALDA,452324,270208,722532,50,6385,7956,6,0,321934,65298,387232,41,5934,327,15,15)  
(West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,560031,210,10149,2882,8,17)  
(West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,322659,918950,73,13452,2787,0,0)  
(West Bengal,MURSHIDABAD,702442,506963,1209405,50,10260,7012,18,0,498998,198174,697172,47,7838,2423,26,26)  
(West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,28,6635,3961,17,41)  
(West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,226104,584064,66,10931,3150,101,0)  
(West Bengal,PURULIA,210168,306933,517101,50,7542,4047,10,0,97160,79169,176329,10,4692,1128,20,0)  
(West Bengal,SILIGURI,59536,25377,84913,30,935,1393,0,10,37794,18060,55854,30,929,906,5,7)  
(West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,162487,756199,31,7257,1631,29,29)  
(West Bengal,UTTAR DINAJPUR,257662,301645,559307,50,4806,1556,30,0,148802,180619,329421,30,2562,2041,17,0)  
grunt> C = FILTER B by Project_Objectives_IHHL_BPL == Project_Performance_IHHL_BPL;  
grunt> STORE C INTO 'hdfs://localhost:8020/PhysicalProgress_BPL_Achieved/ ' USING PigStorage (' ');
```

```
cloudera@quickstart:~  
cloudera:8020/flume_import/FlumeData.1509373363143  
-rw-r--r-- 1 cloudera supergroup 966 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363144  
-rw-r--r-- 1 cloudera supergroup 846 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363145  
-rw-r--r-- 1 cloudera supergroup 967 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363146  
-rw-r--r-- 1 cloudera supergroup 834 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363147  
-rw-r--r-- 1 cloudera supergroup 963 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363148  
-rw-r--r-- 1 cloudera supergroup 831 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363149  
-rw-r--r-- 1 cloudera supergroup 959 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363150  
-rw-r--r-- 1 cloudera supergroup 846 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363151  
-rw-r--r-- 1 cloudera supergroup 966 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363152  
-rw-r--r-- 1 cloudera supergroup 841 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363153  
-rw-r--r-- 1 cloudera supergroup 965 2017-10-30 07:23 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363154  
-rw-r--r-- 1 cloudera supergroup 182 2017-10-30 07:24 hdfs://quickstart  
cloudera:8020/flume_import/FlumeData.1509373363155  
[cloudera@quickstart ~]$ hadoop fs -ls hdfs://quickstart.cloudera:8020/  
Found 12 items  
drwxr-xr-x - cloudera supergroup 0 2017-10-28 04:24 hdfs://quickstart.cloudera:8020/PhysicalProgressAllRecords  
drwxr-xr-x - cloudera supergroup 0 2017-10-27 12:58 hdfs://quickstart.cloudera:8020/PhysicalProgress_BPL_80%Reached  
drwxr-xr-x - cloudera supergroup 0 2017-10-26 12:24 hdfs://quickstart.cloudera:8020/PhysicalProgress_BPL_Achieved  
drwxrwxrwx - hdfs supergroup 0 2017-09-19 06:29 hdfs://quickstart.cloudera:8020/benchmarks  
drwxr-xr-x - cloudera supergroup 0 2017-10-30 07:24 hdfs://quickstart.cloudera:8020/flume_import  
drwxr-xr-x - hbase supergroup 0 2017-10-30 03:56 hdfs://quickstart.cloudera:8020/hbase  
drwxr-xr-x - solr solr 0 2017-07-19 06:31 hdfs://quickstart.cloudera:8020/solr  
drwxr-xr-x - cloudera supergroup 0 2017-10-21 05:12 hdfs://quickstart.cloudera:8020/sqoopout1  
drwxr-xr-x - cloudera supergroup 0 2017-10-21 05:32 hdfs://quickstart.cloudera:8020/sqoopout2  
drwxrwxrwt - hdfs supergroup 0 2017-10-25 05:33 hdfs://quickstart.cloudera:8020/tmp  
drwxr-xr-x - hdfs supergroup 0 2017-07-19 06:31 hdfs://quickstart.cloudera:8020/user  
drwxr-xr-x - hdfs supergroup 0 2017-07-19 06:31 hdfs://quickstart.cloudera:8020/var  
[cloudera@quickstart ~]$ hadoop fs -ls hdfs://quickstart.cloudera:8020/PhysicalProgress_BPL_Achieved  
Found 2 items  
-rw-r--r-- 3 cloudera supergroup 0 2017-10-26 12:24 hdfs://quickstart.cloudera:8020/PhysicalProgress_BPL_Achieved/_SUCCESS  
-rw-r--r-- 3 cloudera supergroup 6352 2017-10-26 12:24 hdfs://quickstart.cloudera:8020/PhysicalProgress_BPL_Achieved/part-m-00000  
[cloudera@quickstart ~]$
```

```
cloudera@quickstart~  
[cloudera@quickstart ~]$ sudo service mysql start  
Starting mysqld: [ OK ]  
[cloudera@quickstart ~]$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 30  
Server version: 5.1.73 Source distribution  
  
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
  
mysql> create table miniproj.PhysicalProgress_BPL_Achieved  
-> (  
-> State_Name varchar(50),  
-> District_Name varchar(50),  
-> Project_Objectives_IHHL_BPL varchar(50),  
-> Project_Objectives_IHHL_APL varchar(50),  
-> Project_Objectives_IHHL_TOTAL varchar(50),  
-> Project_Objectives_SCW varchar(50),  
-> Project_Objectives_School_Toilets varchar(50),  
-> Project_Objectives_Anganwadi_Toilets varchar(50),  
-> Project_Objectives_RSM varchar(50),  
-> Project_Objectives_PC varchar(50),  
-> Project_Performance_IHHL_BPL varchar(50),  
-> Project_Performance_IHHL_APL varchar(50),  
-> Project_Performance_IHHL_TOTAL varchar(50),  
-> Project_Performance_SCW varchar(50),  
-> Project_Performance_School_Toilets varchar(50),  
-> Project_Performance_Anganwadi_Toilets varchar(50),  
-> Project_Performance_RSM varchar(50),  
-> Project_Performance_PC varchar(50),  
-> );  
Query OK, 0 rows affected (0.08 sec)  
  
mysql>
```

```
cloudera@quickstart~  
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_Achieved' --export-dir '/PhysicalProg  
ress_BPL_Achieved' --input-fields-terminated-by ',' -m 1 --columns State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_I  
HHL_TOTAL,Project_Objectives_SCW,Project_Objectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance  
_IHHL_BPL,Project_Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_Toilets,P  
roject_Performance_RSM,Project_Performance_PC  
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
17/10/30 07:41:55 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0  
Enter password:
```

Search the web and Windows

8:12 PM
10/30/2017


```
cloudera@quickstart:~  
grunt> REGISTER 'piggybank-0.15.0.jar';  
grunt> A = LOAD 'StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig  
.piggybank.storage.XMLLoader('row') AS (x:chararray);
```

After dump A, below is the result:

```
cloudera@quickstart:~  
>> XPath (x, 'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,  
>> XPath (x, 'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_T  
OTAL,  
>> XPath (x, 'row/Project_Objectives_SCW') AS Project_Objectives_SCW,  
>> XPath (x, 'row/Project_Objectives_School_Toilets') AS Project_Objectives_Scho  
ol_Toilets,  
>> XPath (x, 'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_A  
nganwadi_Toilets,  
>> XPath (x, 'row/Project_Objectives_RSM') AS Project_Objectives_RSM,  
>> XPath (x, 'row/Project_Objectives_PC') AS Project_Objectives_PC,  
>> XPath (x, 'row/Project_Performance_IHHL_BPL') AS Project_Performance_IHHL_BPL  
,  
>> XPath (x, 'row/Project_Performance_IHHL_APL') AS Project_Performance_IHHL_APL  
,  
>> XPath (x, 'row/Project_Performance_IHHL_TOTAL') AS Project_Performance_IHHL_T  
OTAL,  
>> XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,  
>> XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_Sc  
hool_Toilets,  
>> XPath (x, 'row/Project_Performance_Anganwadi_Toilets') AS Project_Performance  
_Anganwadi_Toilets,  
>> XPath (x, 'row/Project_Performance_RSM') AS Project_Performance_RSM,  
>> XPath (x, 'row/Project_Performance_PC') AS Project_Performance_PC;  
grunt> Dump B;
```

After dump B, below is the result:

```
cloudera@quickstart:~  
42,5178,2733,14,20)  
(West Bengal,JALPAIGURI,372999,203523,576522,50,6578,5428,87,0,337740,101550,439  
290,25,6578,4064,17,14)  
(West Bengal,MALDA,452324,270208,722532,50,6385,7956,6,0,321934,65298,387232,41,  
5934,327,15,15)  
(West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,5  
60031,210,10149,2882,8,17)  
(West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,322659  
,918950,73,13452,2787,0,0)  
(West Bengal,MURSHIDABAD,702442,506963,1209405,50,10260,7012,18,0,498998,198174,  
697172,47,7838,2423,26,26)  
(West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,2  
8,6635,3961,17,41)  
(West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,2  
26104,584064,66,10931,3150,101,0)  
(West Bengal,PURULIA,210168,306933,517101,50,7542,4047,10,0,97160,79169,176329,1  
0,4692,1128,20,0)  
(West Bengal,SILIGURI,59536,25377,84913,30,935,1393,0,10,37794,18060,55854,30,92  
9,906,5,7)  
(West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,1  
62487,756199,31,7257,1631,29,29)  
(West Bengal,UTTAR DINAJPUR,257662,301645,559307,50,4806,1556,30,0,148802,180619  
,329421,30,2562,2041,17,0)  
grunt>
```

After Executing, register One.jar;

D = FOREACH B GENERATE pigudf.FilterBPL1(

State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_Objectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project_Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_Toilets,Project_Performance_RSM,Project_Performance_PC

);

```
cloudera@quickstart:~  
697172,47,7838,2423,26,26)  
(West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,2  
8,6635,3961,17,41)  
(West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,2  
26104,584064,66,10931,3150,101,0)  
(West Bengal,PURULIA,210168,306933,517101,50,7542,4047,10,0,97160,79169,176329,1  
0,4692,1128,20,0)  
(West Bengal,SILIGURI,59536,25377,84913,30,935,1393,0,10,37794,18060,55854,30,92  
9,906,5,7)  
(West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,1  
62487,756199,31,7257,1631,29,29)  
(West Bengal,UTTAR DINAJPUR,257662,301645,559307,50,4806,1556,30,0,148802,180619  
,329421,30,2562,2041,17,0)  
grunt> register One.jar;  
grunt> D = FOREACH B GENERATE pigudf.FilterBPL1(  
>> State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_  
APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_Objectives_Scho  
ol_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_O  
bjectives_PC,Project_Performance_IHHL_BPL,Project_Performance_IHHL_APL,Project_P  
erformance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets  
,Project_Performance_Anganwadi_Toilets,Project_Performance_RSM,Project_Performan  
ce_PC  
>> );  
grunt>
```

After dump D, below is the result:

```
cloudera@quickstart:~  
48,233601,19,2632,939,8,8))  
(  
( (West Bengal,HOOGHLY,271737,195510,467247,53,6821,4168,19,0,269779,191294,46107  
3,49,6764,3435,18,18))  
( (West Bengal,HOWRAH,231860,143309,375169,51,5195,3586,26,0,230190,141912,372102  
,42,5178,2733,14,20))  
( (West Bengal,JALPAIGURI,372999,203523,576522,50,6578,5428,87,0,337740,101550,43  
9290,25,6578,4064,17,14))  
(  
( (West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,  
560031,210,10149,2882,8,17))  
( (West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,32265  
9,918950,73,13452,2787,0,0))  
(  
( (West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,  
28,6635,3961,17,41))  
( (West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,  
226104,584064,66,10931,3150,101,0))  
(  
(  
( (West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,  
162487,756199,31,7257,1631,29,29))  
(  
grunt>
```

After executing the below command, below is the screen:

D_CLEAN = FILTER D BY NOT (\$0 IS NULL);

DUMP D_CLEAN;

```
cloudera@quickstart:~  
(  
( (West Bengal,HOOGLHY,271737,195510,467247,53,6821,4168,19,0,269779,191294,46107  
3,49,6764,3435,18,18) )  
( (West Bengal,HOWRAH,231860,143309,375169,51,5195,3586,26,0,230190,141912,372102  
,42,5178,2733,14,20) )  
( (West Bengal,JALPAIGURI,372999,203523,576522,50,6578,5428,87,0,337740,101550,43  
9290,25,6578,4064,17,14) )  
(  
( (West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,  
560031,210,10149,2882,8,17) )  
( (West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,32265  
9,918950,73,13452,2787,0,0) )  
(  
( (West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,  
28,6635,3961,17,41) )  
( (West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,  
226104,584064,66,10931,3150,101,0) )  
(  
(  
( (West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,  
162487,756199,31,7257,1631,29,29) )  
(  
grunt> D_CLEAN = FILTER D BY NOT ($0 IS NULL);  
grunt> DUMP D_CLEAN;
```

Acadgild - Mini Project Report - Word

File Home Insert Design Layout References Mailings Review View Tell me what you want to do

cloudera@quickstart:~

```
88,0,119,75,4,1))  
( (Uttarakhand,UTTARKASHI,28189,20700,48889,50,240,123,4,0,25523,21017,46540,4,24  
0,0,3,0) )  
( (West Bengal,BARDHAMAN,700047,341920,1041967,133,9891,7980,10,0,601906,277914,8  
79620,140,9890,7724,10,19) )  
( (West Bengal,DAKHIN DINAJPUR,182621,194577,377198,50,3712,2642,10,0,184153,494  
48,233601,19,2632,939,8,8) )  
( (West Bengal,HOOGLHY,271737,195510,467247,53,6821,4168,19,0,269779,191294,46107  
3,49,6764,3435,18,18) )  
( (West Bengal,HOWRAH,231860,143309,375169,51,5195,3586,26,0,230190,141912,372102  
,42,5178,2733,14,20) )  
( (West Bengal,JALPAIGURI,372999,203523,576522,50,6578,5428,87,0,337740,101550,43  
9290,25,6578,4064,17,14) )  
( (West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,  
560031,210,10149,2882,8,17) )  
( (West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,32265  
9,918950,73,13452,2787,0,0) )  
( (West Bengal,NADIA,346696,278335,625031,50,6974,6620,50,0,321462,198890,520352,  
28,6635,3961,17,41) )  
( (West Bengal,NORTH 24 PARAGANAS,361462,225080,586542,51,11158,4466,30,0,357960,  
226104,584064,66,10931,3150,101,0) )  
( (West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,  
162487,756199,31,7257,1631,29,29) )  
grunt>  
(  
(  
( (West Bengal,SOUTH 24 PARAGANAS,628712,521192,1149904,50,8940,5448,30,0,593712,  
162487,756199,31,7257,1631,29,29) )  
(  
grunt> D_CLEAN = FILTER D BY NOT ($0 IS NULL);  
grunt> DUMP D_CLEAN;
```

Page 15 of 16 524 words

Search the web and Windows

8:54 PM 10/30/2017

After executing- STORE D_CLEAN INTO 'hdfs://localhost:8020/PhysicalProgress_BPL_80%Reached/'
USING PigStorage (',');

```
cloudera@quickstart:~  
017-10-30 08:29:40      FILTER  
Success!  
  
Job Stats (time in seconds):  
JobId  Alias  Feature Outputs  
job_local1676275090_0007      A,B,D,D_CLEAN  MAP_ONLY      hdfs://localhost  
:8020/PhysicalProgress_BPL_80PercentReached,  
  
Input(s):  
Successfully read records from: "file:///home/cloudera/StatewiseDistrictwisePhysicalProgress.xml"  
  
Output(s):  
Successfully stored records in: "hdfs://localhost:8020/PhysicalProgress_BPL_80PercentReached"  
  
Job DAG:  
job_local1676275090_0007  
  
2017-10-30 08:29:46,715 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
grunt>
```

After following command in MYSQL, below is the result:

```
cloudera@quickstart:~  
-----+-----  
70 rows in set (0.00 sec)  
  
mysql> desc miniproj.PhysicalProgress_BPL_80PercentReached;  
ERROR 1146 (42S02): Table 'miniproj.PhysicalProgress_BPL_80PercentReached' doesn't exist  
mysql> show tables;  
ERROR 1046 (3D000): No database selected  
mysql> use miniproj;  
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> show tables;  
+-----+-----+  
| Tables_in_miniproj |  
+-----+-----+  
| PhysicalProgress_BPL_Achieved |  
+-----+-----+  
1 row in set (0.00 sec)  
  
mysql> create table miniproj.PhysicalProgress_BPL_80PercentReached  
-> (  
-> State_Name varchar(50),  
-> District_Name varchar(50),  
-> Project_Objectives_IHHL_BPL varchar(50),  
-> Project_Objectives_IHHL_APL varchar(50),  
-> Project_Objectives_IHHL_TOTAL varchar(50),  
-> Project_Objectives_SCW varchar(50),  
-> Project_Objectives_School_Toilets varchar(50),  
-> Project_Objectives_Anganwadi_Toilets varchar(50),  
-> Project_Objectives_RSM varchar(50),  
-> Project_Objectives_PC varchar(50),  
-> Project_Performance_IHHL_BPL varchar(50),  
-> Project_Performance_IHHL_APL varchar(50),  
-> Project_Performance_IHHL_TOTAL varchar(50),  
-> Project_Performance_SCW varchar(50),  
-> Project_Performance_School_Toilets varchar(50),  
-> Project_Performance_Anganwadi_Toilets varchar(50),  
-> Project_Performance_RSM varchar(50),  
-> Project_Performance_PC varchar(50)  
-> );  
Query OK, 0 rows affected (0.07 sec)  
  
mysql>
```

In following screen, below command is executed:

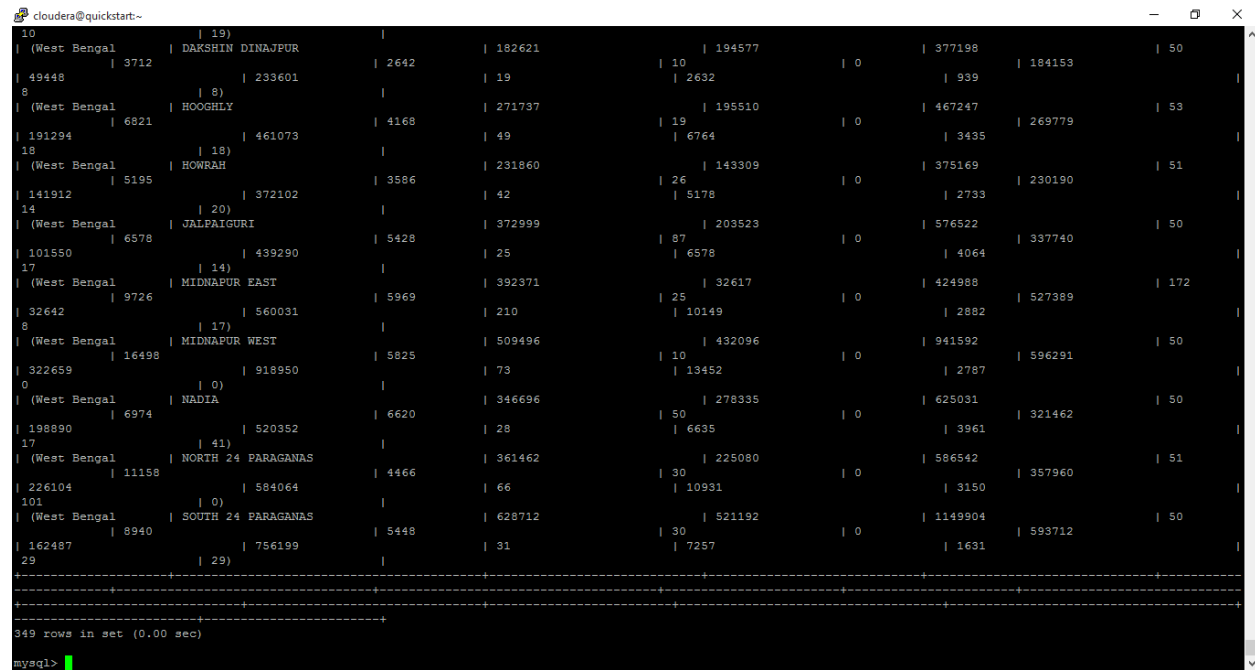
```
cloudera@quickstart:~$ sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir '/PhysicalProgress_BPL_80PercentReached' --input-fields-terminated-by ',' -m 1 --columns State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_Objectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project_Performance_IHHL_APL,Project_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_Anganwadi_Toilets,Project_Performance_RSM,Project_Performance_PC;
```

After executing the Linux command, show the below screen shot:

```
sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table  
'PhysicalProgress_BPL_80PercentReached' --export-dir '/PhysicalProgress_BPL_80PercentReached' --input-fields-  
terminated-by ',' -m 1 --columns  
State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_  
TOTAL,Project_Objectives_SCW,Project_Objectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_  
_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project_Performance_IHHL_APL,Project_  
_Performance_IHHL_TOTAL,Project_Performance_SCW,Project_Performance_School_Toilets,Project_Performance_  
_Anganwadi_Toilets,Project_Performance_RSM,Project_Performance_PC;
```

```
cloudera@quickstart:~$  
17/10/30 08:38:33 INFO mapreduce.Job: Job job_1509360947720_0002 running in uber mode : false  
17/10/30 08:38:33 INFO mapreduce.Job: map 0% reduce 0%  
17/10/30 08:38:36 INFO mapreduce.Job: map 100% reduce 0%  
17/10/30 08:38:58 INFO mapreduce.Job: Job job_1509360947720_0002 completed successfully  
17/10/30 08:38:58 INFO mapreduce.Job: Counters: 30  
  File System Counters  
    FILE: Number of bytes read=0  
    FILE: Number of bytes written=152461  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=32854  
    HDFS: Number of bytes written=0  
    HDFS: Number of read operations=4  
    HDFS: Number of large read operations=0  
    HDFS: Number of write operations=0  
  Job Counters  
    Launched map tasks=1  
    Data-local map tasks=1  
    Total time spent by all maps in occupied slots (ms)=21468  
    Total time spent by all reduces in occupied slots (ms)=0  
    Total time spent by all map tasks (ms)=21468  
    Total vcore-milliseconds taken by all map tasks=21468  
    Total megabyte-milliseconds taken by all map tasks=21983232  
  Map-Reduce Framework  
    Map input records=349  
    Map output records=349  
    Input split bytes=163  
    Spilled Records=0  
    Failed Shuffles=0  
    Merged Map outputs=0  
    GC time elapsed (ms)=129  
    CPU time spent (ms)=1490  
    Physical memory (bytes) snapshot=111972352  
    Virtual memory (bytes) snapshot=1508073472  
    Total committed heap usage (bytes)=60882944  
  File Input Format Counters  
    Bytes Read=0  
  File Output Format Counters  
    Bytes Written=0  
17/10/30 08:38:58 INFO mapreduce.ExportJobBase: Transferred 32.084 KB in 46.7871 seconds (702.2023 bytes/sec)  
17/10/30 08:38:58 INFO mapreduce.ExportJobBase: Exported 349 records.  
cloudera@quickstart:~$
```

After executing the following comment in Mysql, select * from PhysicalProgress_BPL_80PercentReached below is the result:



```
cloudera@quickstart~  
+-----+  
10 | (West Bengal | 19) | | | 182621 | | 194577 | | 377198 | | 50  
| 49448 | 3712 | 233601 | | 19 | 2632 | 0 | 939 | 184153 | |  
8 | (West Bengal | 8) | | | 271737 | 19 | 195510 | | 467247 | 269779 | 53  
| 191294 | 6821 | 461073 | | 49 | 6764 | 0 | 3435 | | |  
18 | (West Bengal | 18) | | | 231860 | | 143309 | | 375169 | 230190 | 51  
| 141912 | 5195 | 372102 | | 42 | 5178 | 0 | 2733 | | |  
14 | (West Bengal | 20) | | | 372999 | | 203523 | | 576522 | 337740 | 50  
| 101550 | 6578 | 439290 | | 25 | 6578 | 0 | 4064 | | |  
17 | (West Bengal | 14) | | | 392371 | 25 | 32617 | | 424988 | 527389 | 172  
| 32642 | 9726 | 560031 | | 210 | 10149 | 0 | 2882 | | |  
8 | (West Bengal | 17) | | | 509496 | | 432096 | | 941592 | 596291 | 50  
| 322659 | 16498 | 918950 | | 73 | 13452 | 0 | 2787 | | |  
0 | (West Bengal | 0) | | | 346696 | | 278335 | | 625031 | 321462 | 50  
| 198890 | 6974 | 520352 | | 28 | 6635 | 0 | 3961 | | |  
17 | (West Bengal | 41) | | | 361462 | 30 | 225080 | | 586542 | 357960 | 51  
| 226104 | 11158 | 584064 | | 66 | 10931 | 0 | 3150 | | |  
101 | (West Bengal | 0) | | | 628712 | | 521192 | | 1149904 | 593712 | 50  
| 162487 | 8940 | 756199 | | 31 | 7257 | 0 | 1631 | | |  
29 | (West Bengal | 29) | | | | | | | | | |  
+-----+  
349 rows in set (0.00 sec)  
mysql>
```

5. Upload the solution document in the GitHub and share the link in the Acadgild dashboard for further evaluation.

Please refer the attached zipped file, to see the code which are used in this implementation.