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

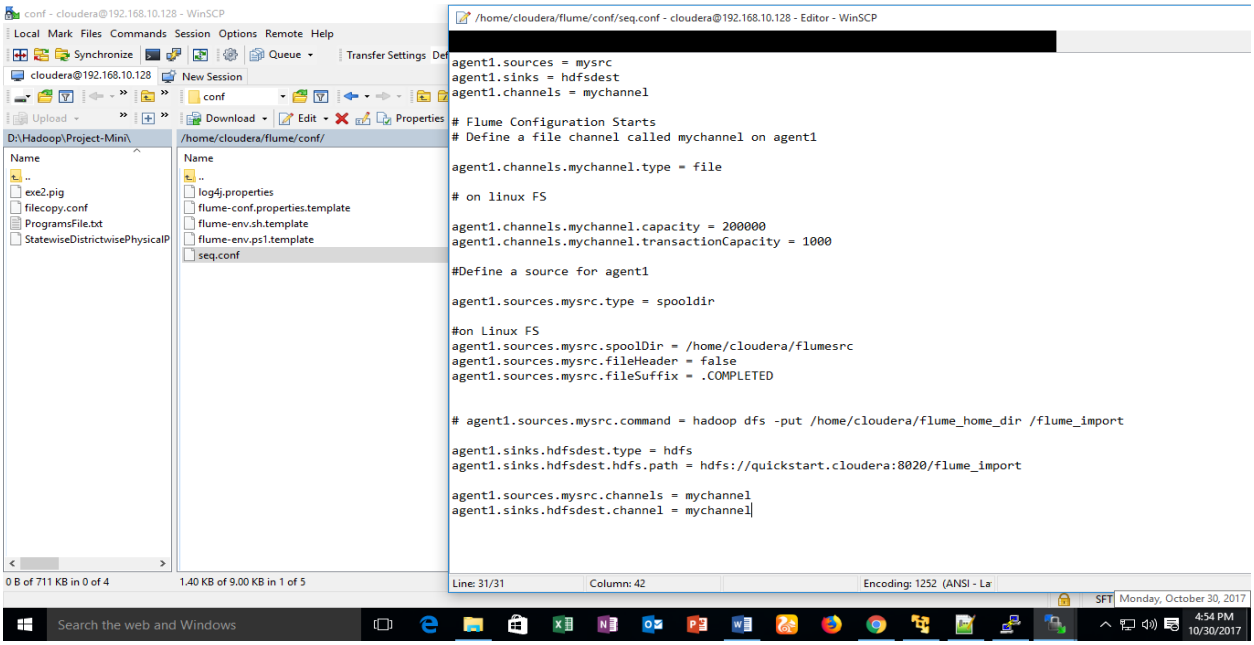
2. Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards.

Export the results to MySQL using Sqoop.

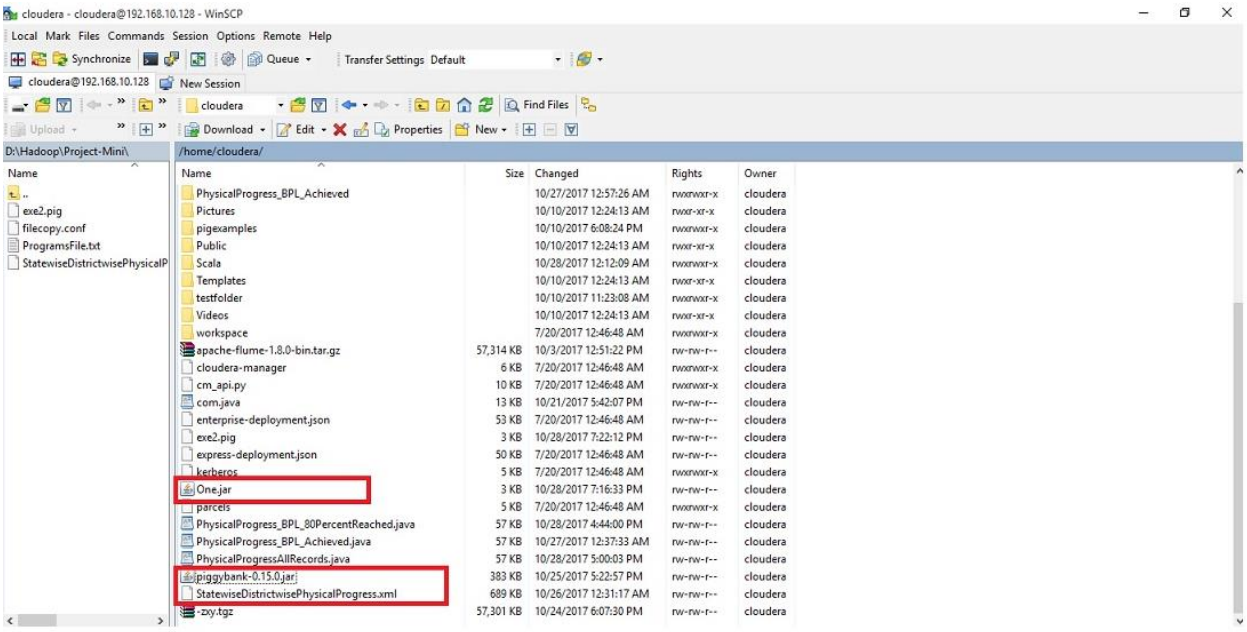
2. Screenshot of the command executed:

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

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

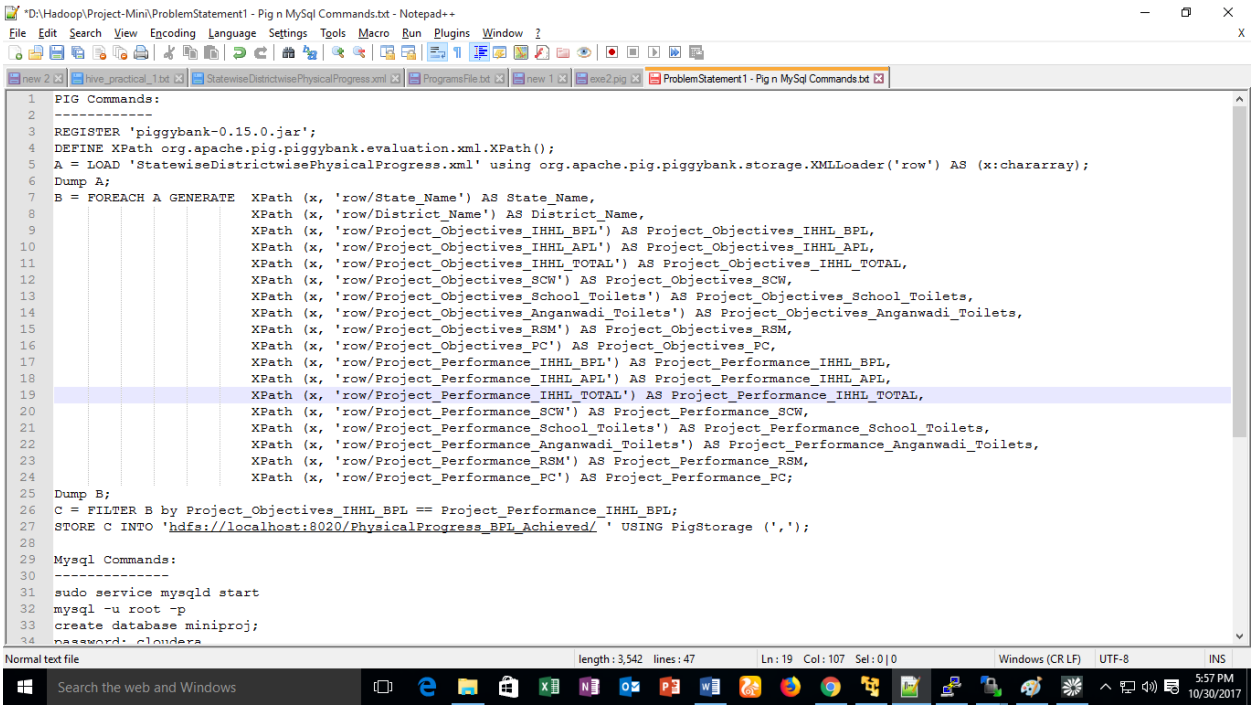


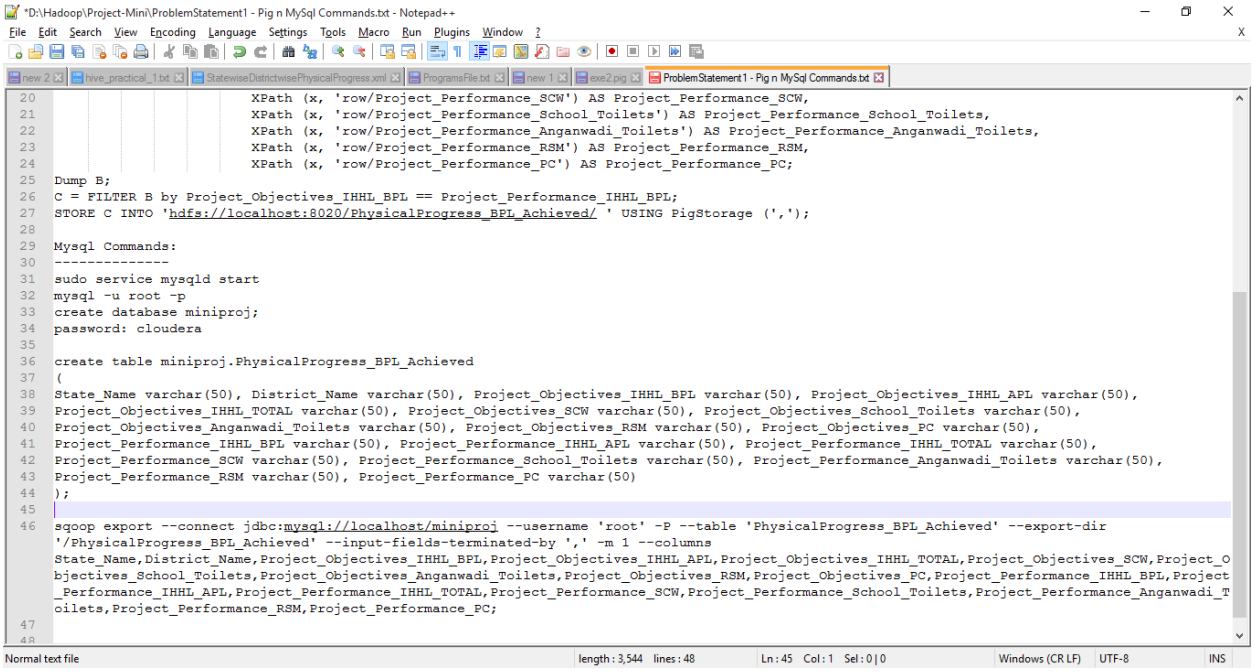
b. Winscp screen shot to depict the files which are moved to the linux directory (Which is common to Project Statement 1 & 2)



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

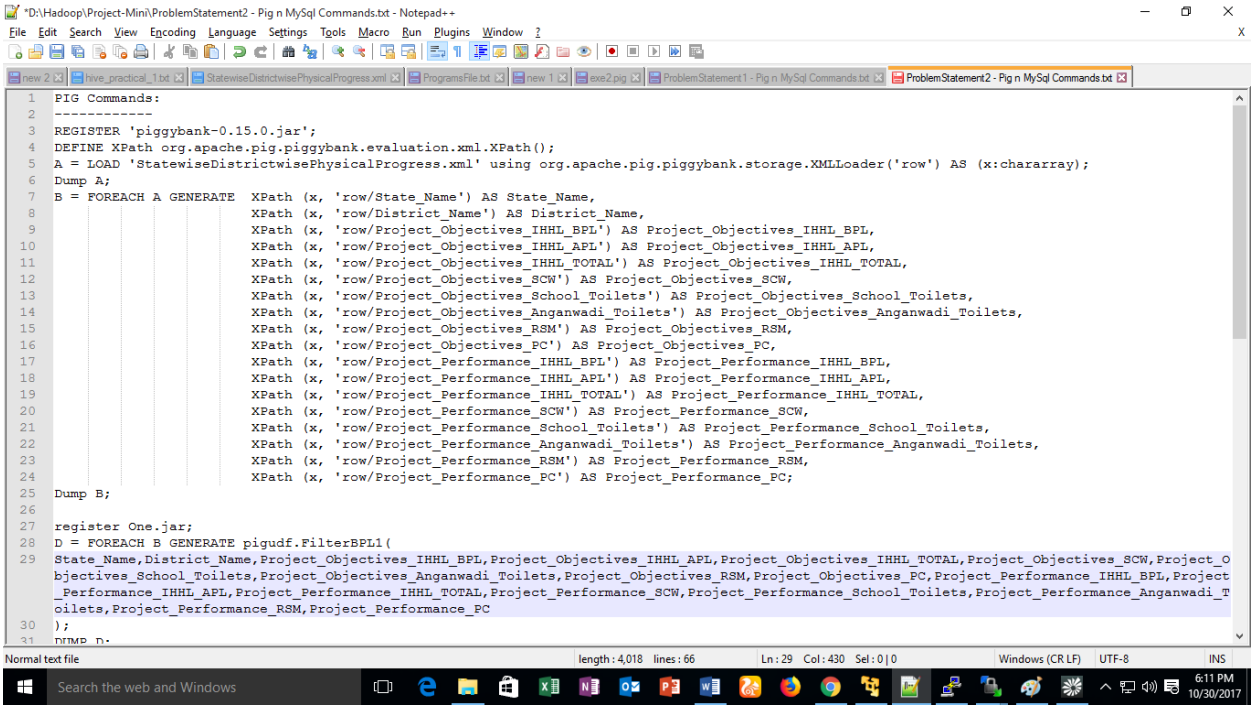
c. Project Statement 1 – Pig and MySql Commands Screen shot:



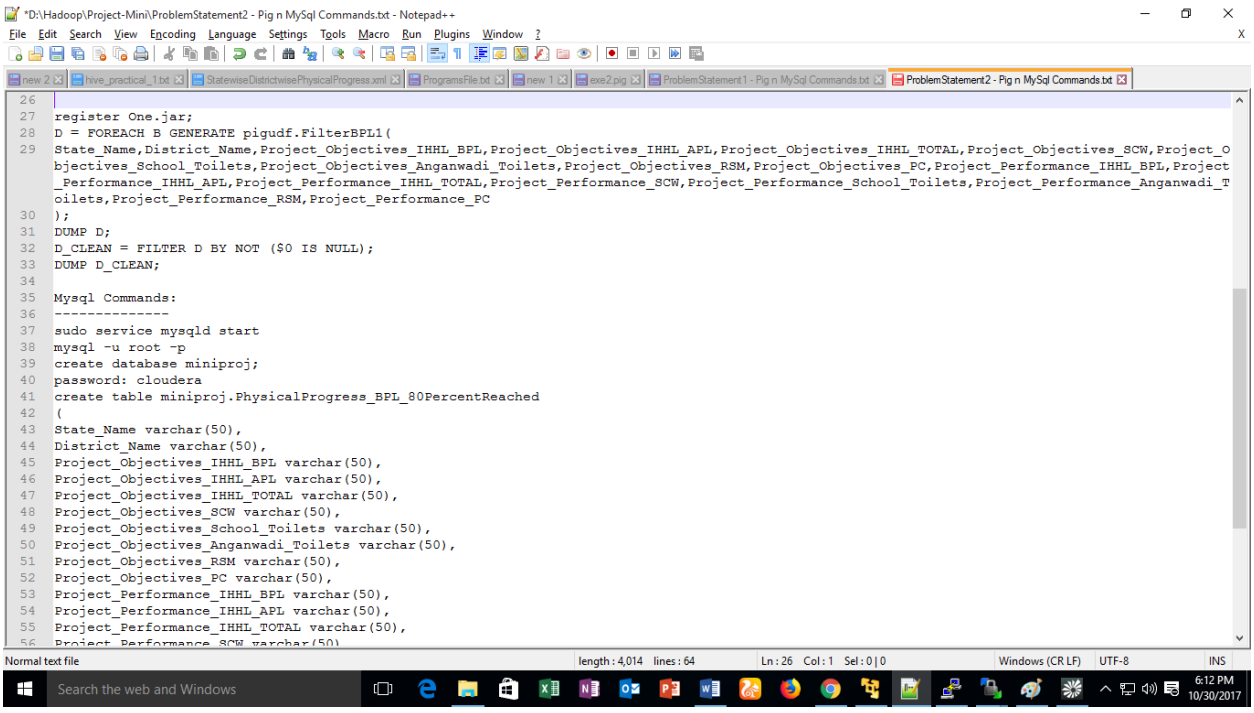


```
                                XPath (x, 'row/Project_Performance_SCW') AS Project_Performance_SCW,
                                XPath (x, 'row/Project_Performance_School_Toilets') AS Project_Performance_School_Toilets,
                                XPath (x, 'row/Project_Performance_Anganwadi_Toilets') 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)
);

sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_Achieved' --export-dir
'/PhysicalProgress_BPL_Achieved' --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;
```

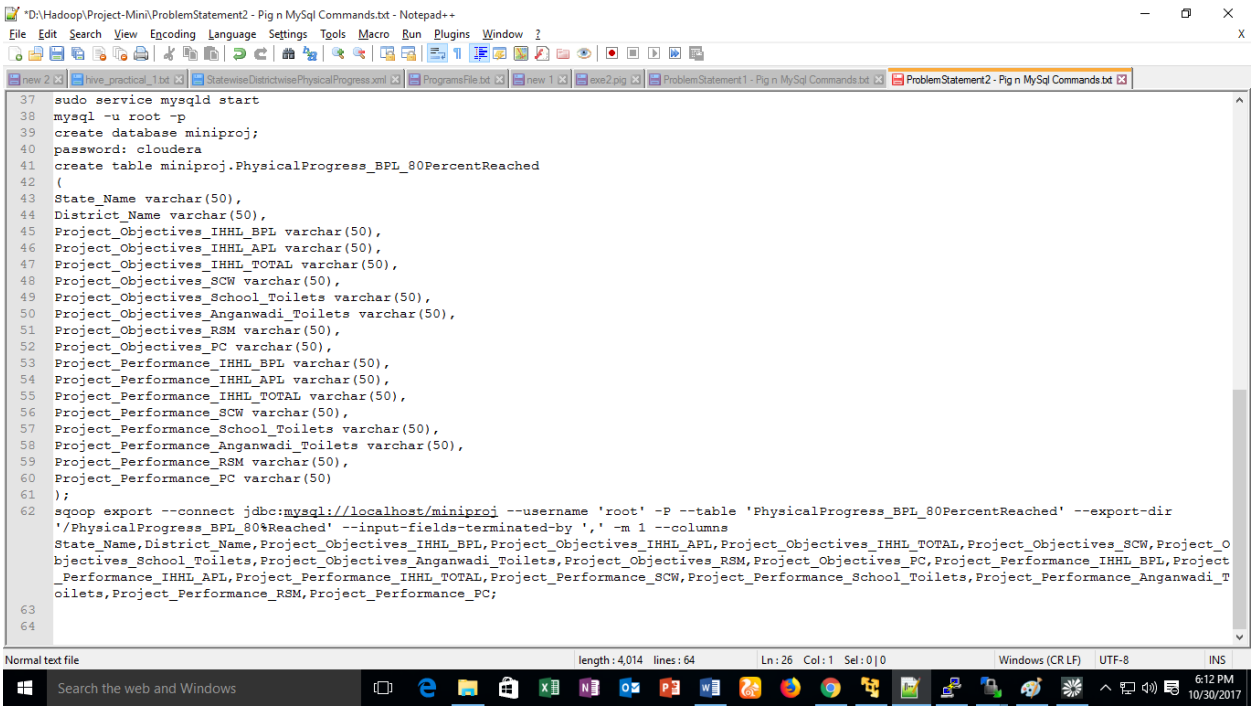
d. Project Statement 2 – Pig and MySql Commands Screen shot:



```
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;

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_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
);
```

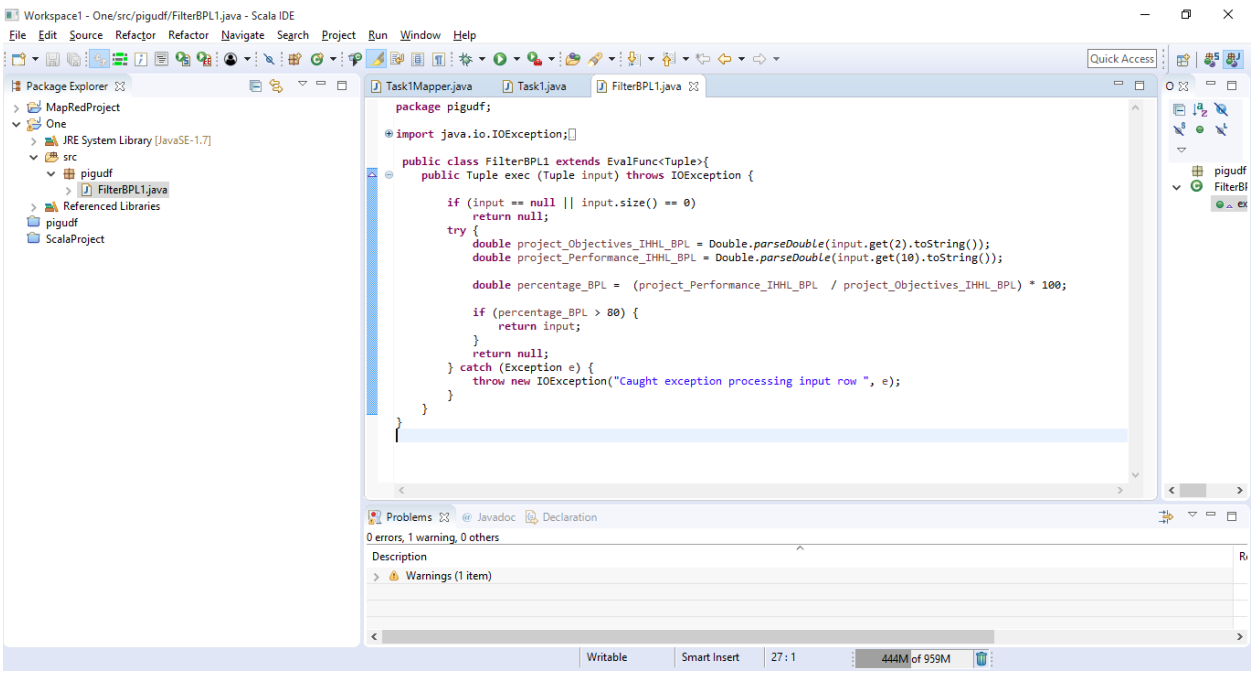


```
register One.jar;
D = FOREACH B GENERATE pigudf.FilterBPL1(
State_Name,District_Name,Project_Objectives_IHHL_BPL,Project_Objectives_IHHL_APL,Project_Objectives_IHHL_TOTAL,Project_Objectives_SCW,Project_O
bjectives_School_Toilets,Project_Objectives_Anganwadi_Toilets,Project_Objectives_RSM,Project_Objectives_PC,Project_Performance_IHHL_BPL,Project
_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
);
DUMP D;
D_CLEAN = FILTER D BY NOT ($0 IS NULL);
DUMP D_CLEAN;

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



```
sudo service mysqld start
mysql -u root -p
create database miniproj;
password: cloudera
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)
);
sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir
'/PhysicalProgress_BPL_80%Reached' --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 commpands
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.

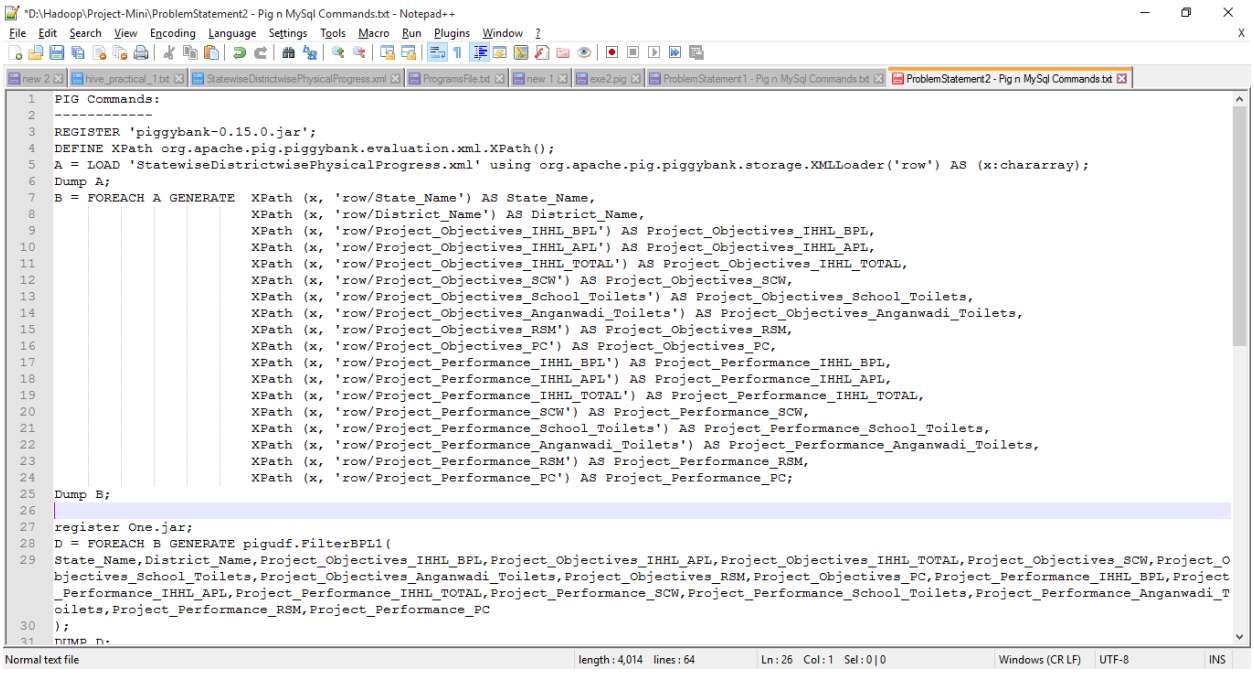
Code Explanation and Way of Working – Common to Problem statement 1. Below screen shot explains the flow and purpose those commands:

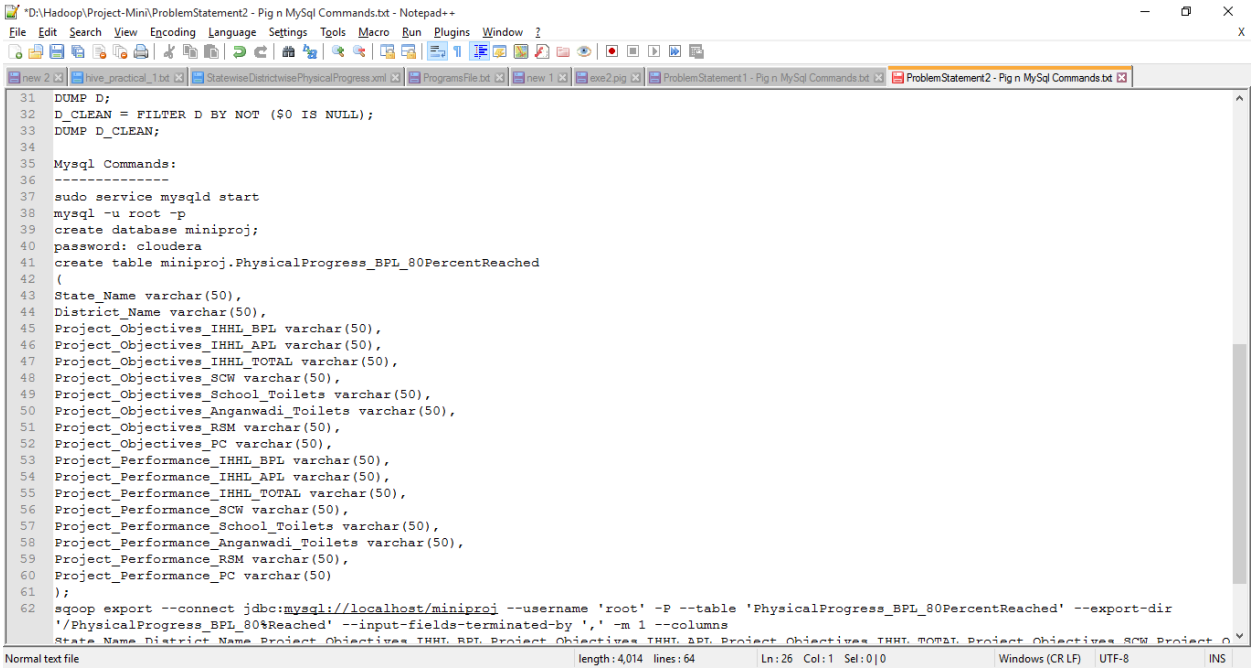
```
PIG Commands:
------------
REGISTER 'piggybank-0.15.0.jar';
DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath();
A = LOAD 'StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') AS (x:chararray);
Dump A;
B = FOREACH A GENERATE  XPath (x, 'row/State_Name') 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:



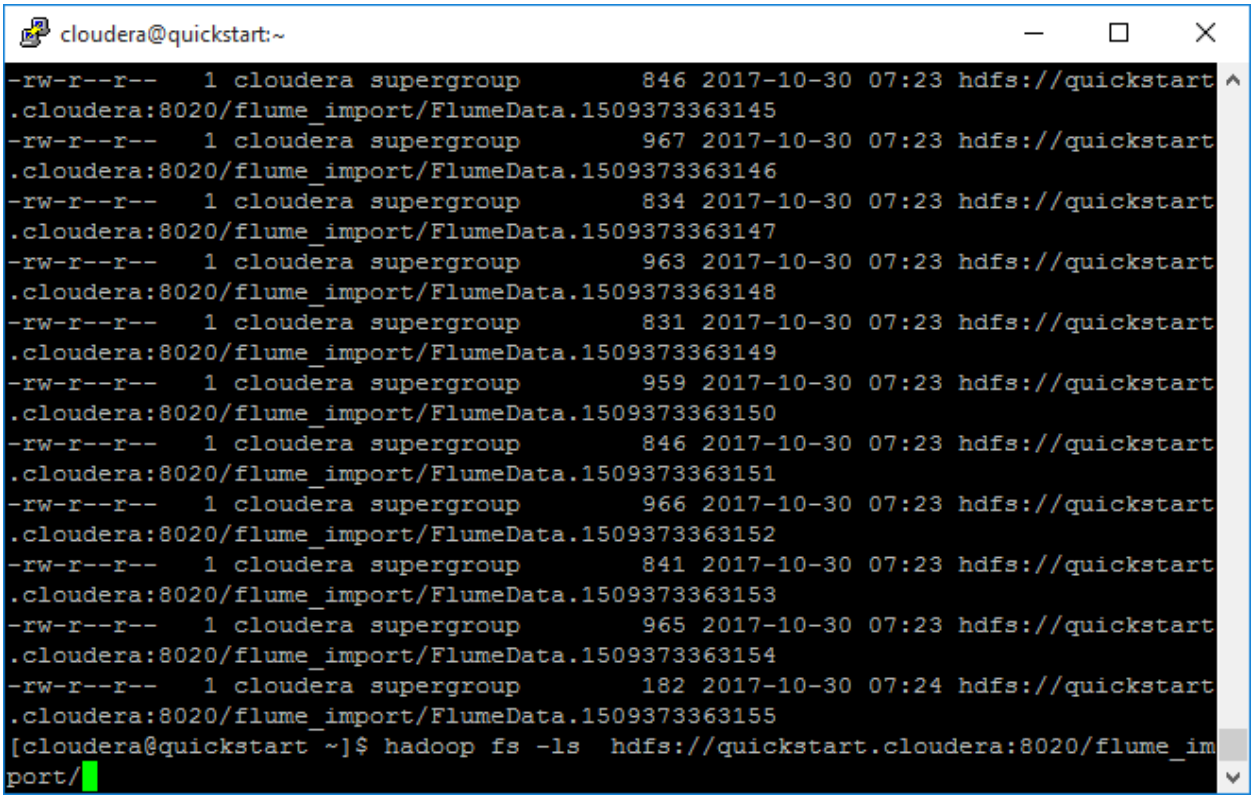


```
DUMP D;
D_CLEAN = FILTER D BY NOT ($0 IS NULL);
DUMP D_CLEAN;

Mysql Commands:
--------------
sudo service mysqld start
mysql -u root -p
create database miniproj;
password: cloudera
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)
);
sqoop export --connect jdbc:mysql://localhost/miniproj --username 'root' -P --table 'PhysicalProgress_BPL_80PercentReached' --export-dir
'/PhysicalProgress_BPL_80%Reached' --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;
```

4. Screenshot of the output

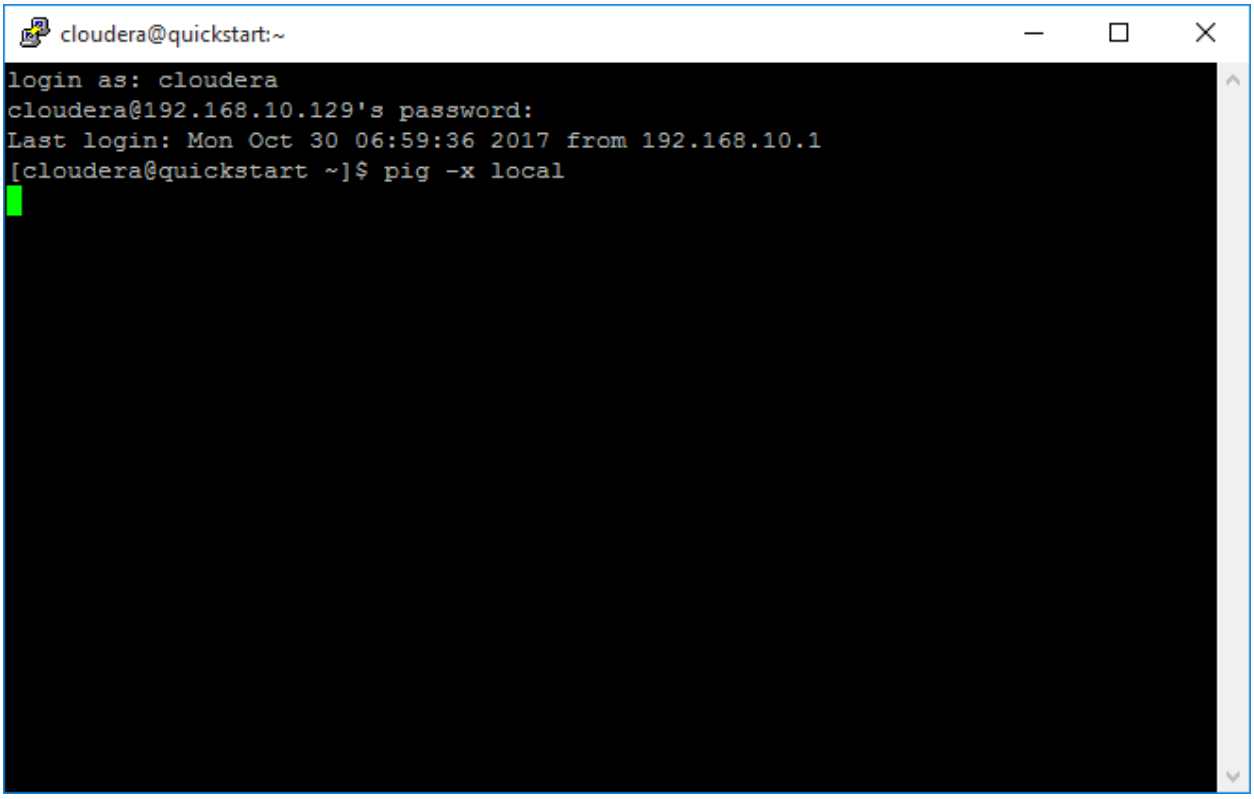
a.

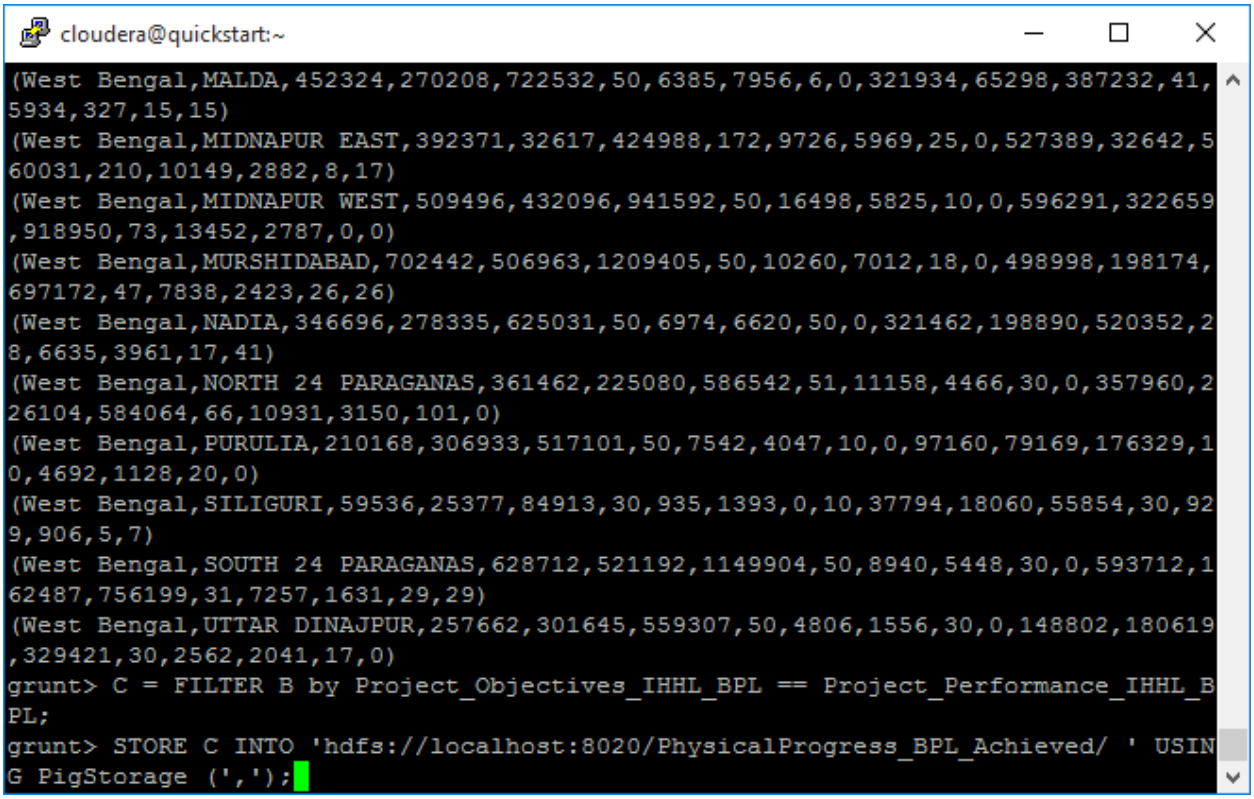
Output of the Flume Job:

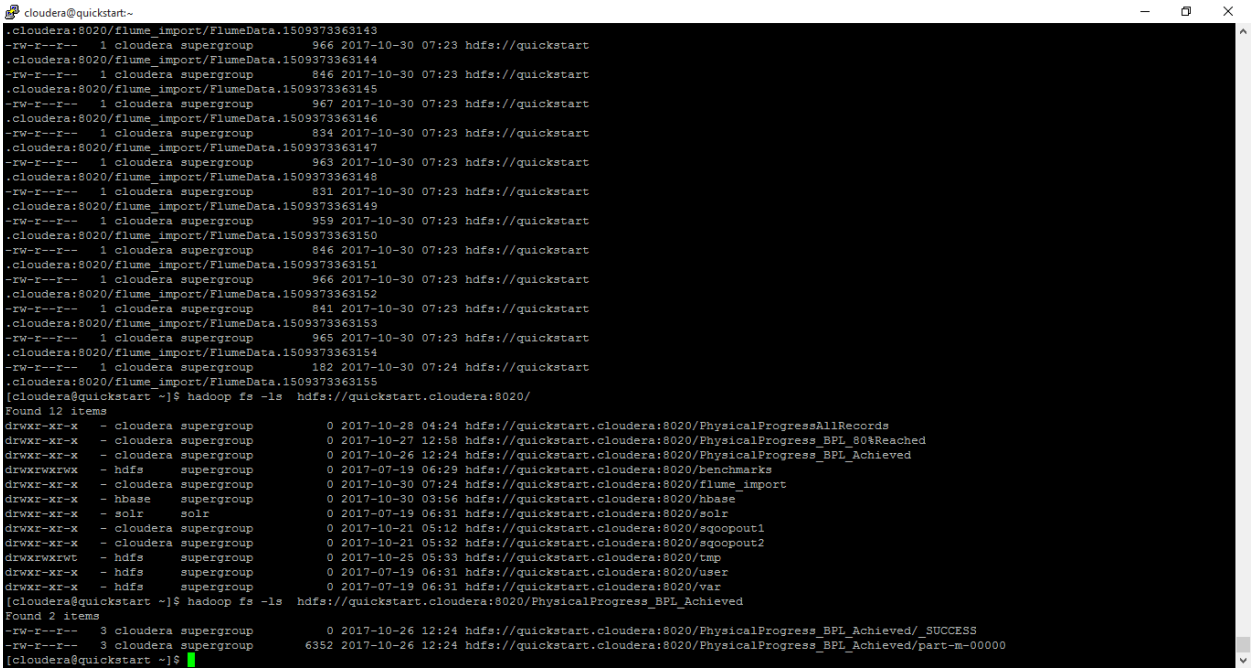


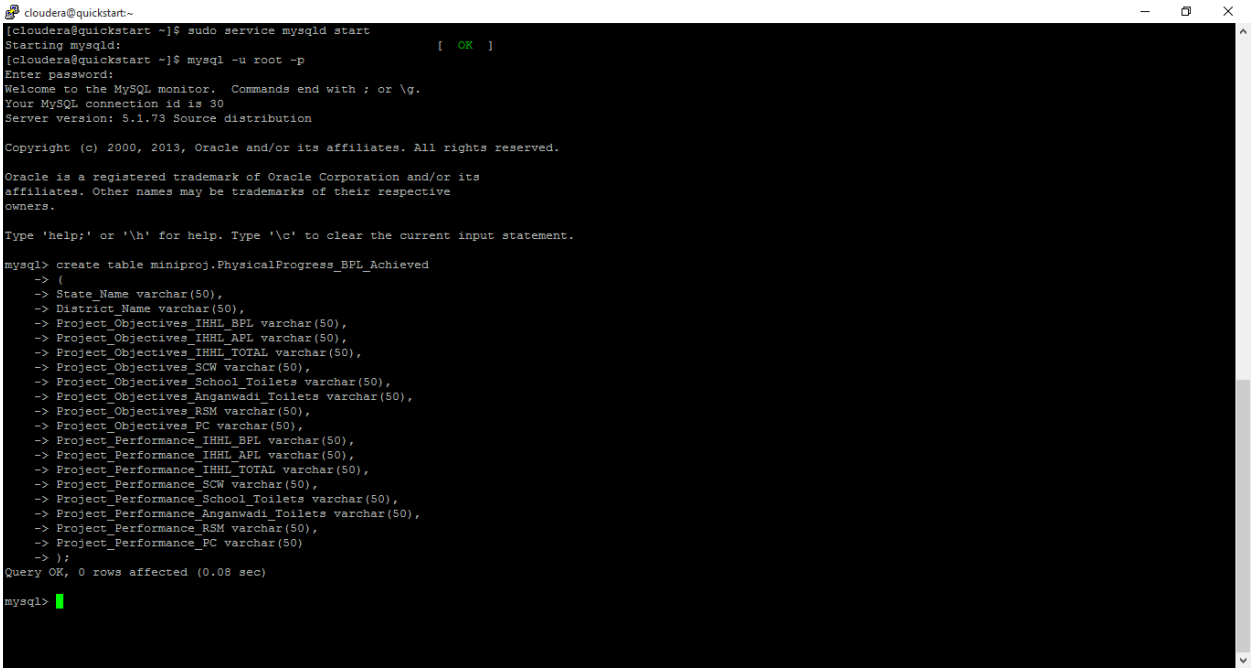
b.

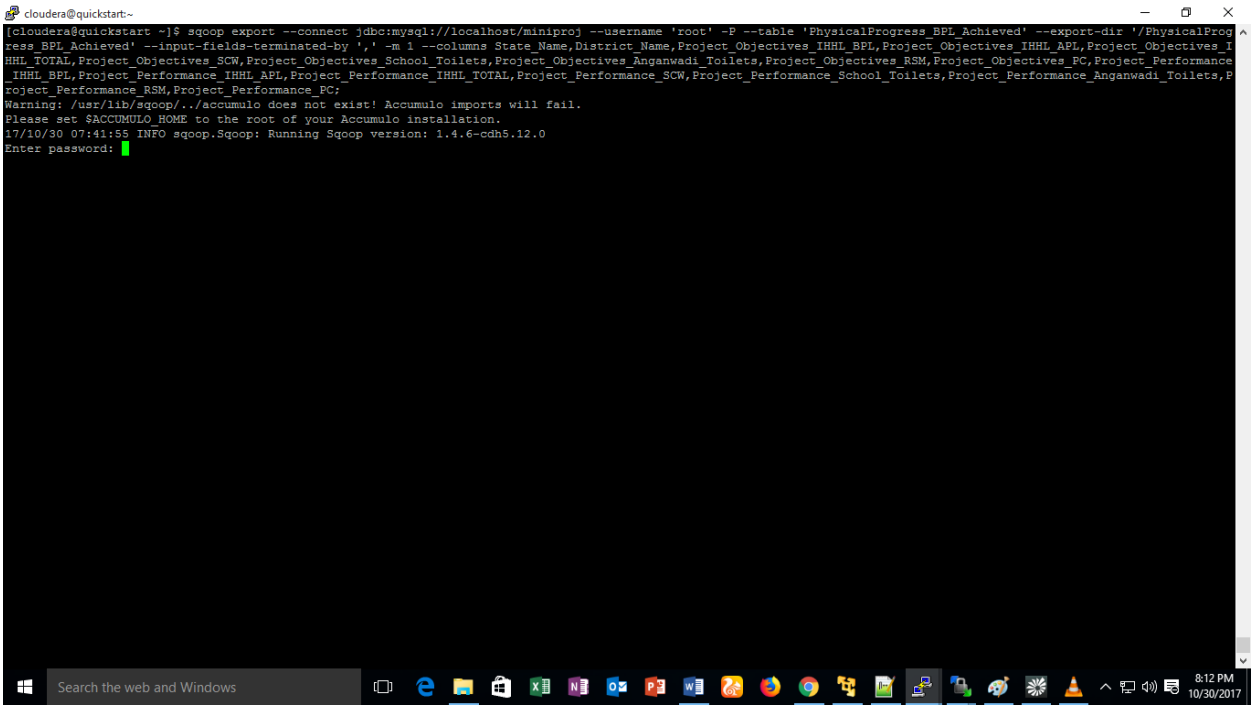
Screen shots for the output of Problem statement 1:

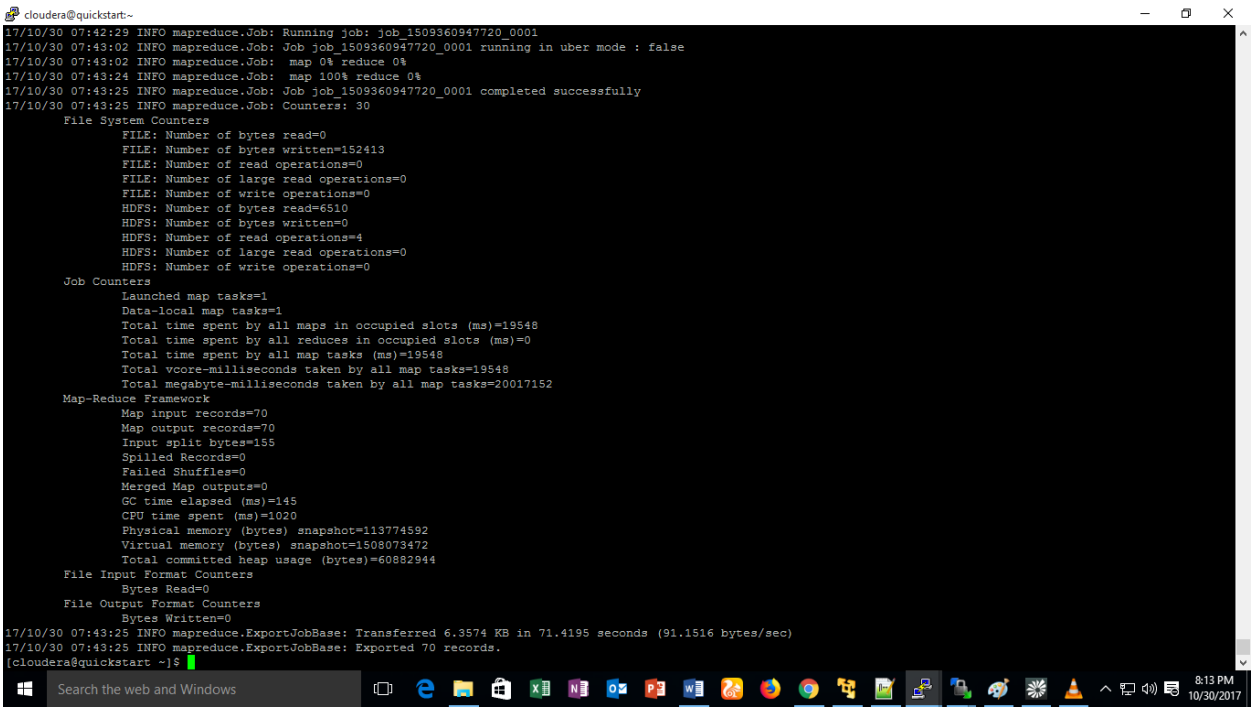


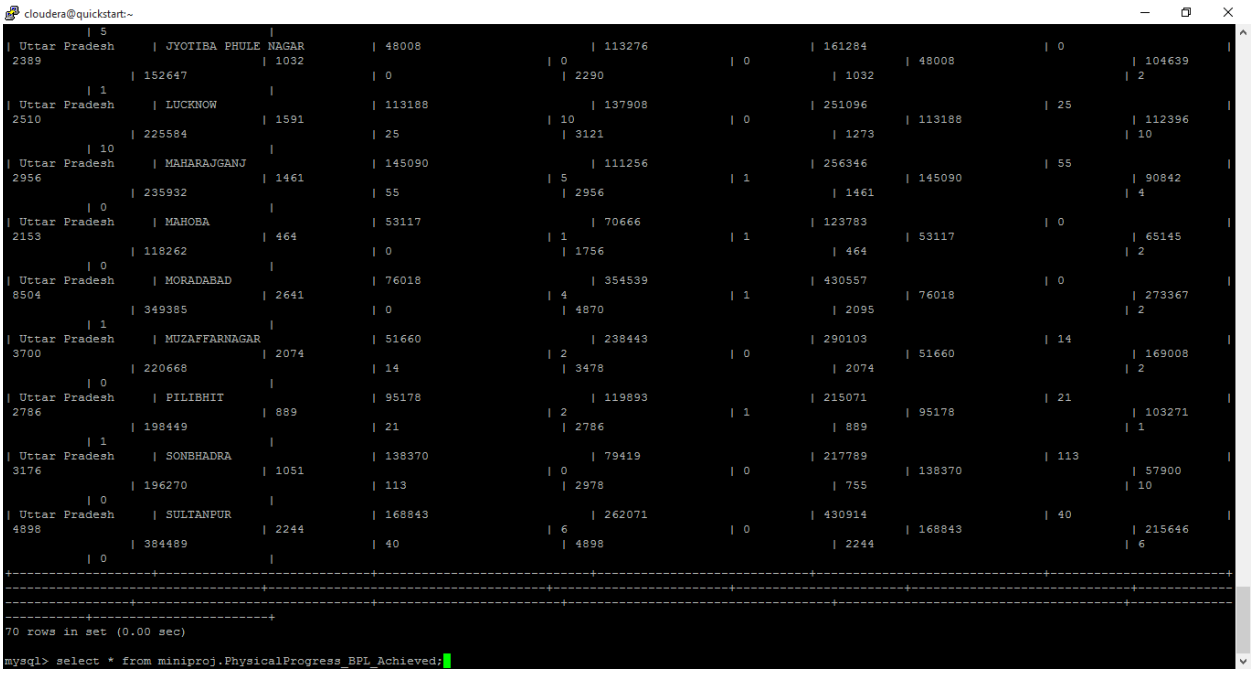

cloudera@quickstart:~
(West Bengal,MALDA,452324,270208,722532,50,6385,7956,6,0,321934,65298,387232,41,5934,327,15,15)
(West Bengal,MIDNAPUR EAST,392371,32617,424988,172,9726,5969,25,0,527389,32642,560031,210,10149,2882,8,17)
(West Bengal,MIDNAPUR WEST,509496,432096,941592,50,16498,5825,10,0,596291,322659,918950,73,13452,2787,0,0)
(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 2017-07-19 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 ~]$

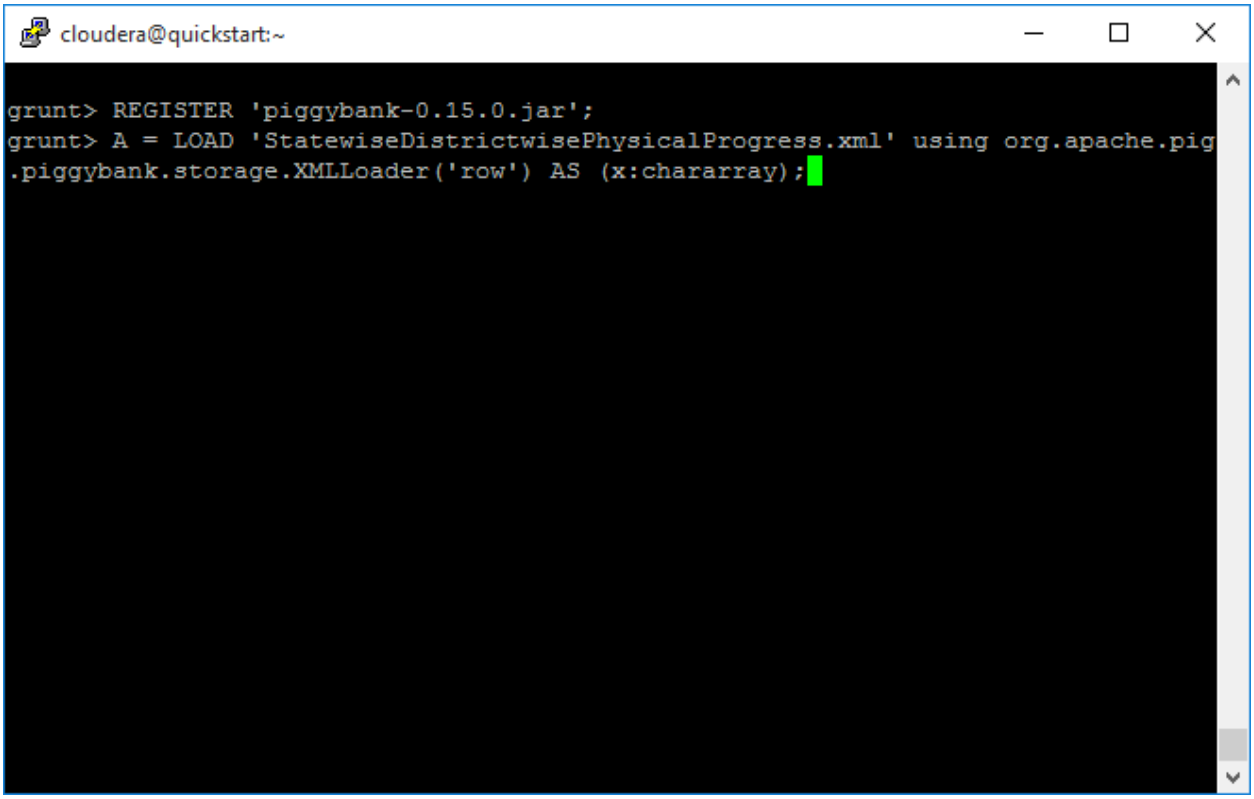

cloudera@quickstart:~
[cloudera@quickstart ~]$ sudo service mysqld start
Starting mysqld: [OK]
[cloudera@quickstart ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 30
Server version: 5.1.73 Source distribution
Copyright (c) 2000, 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 '/PhysicalProgress_BPL_Achieved' --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;
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

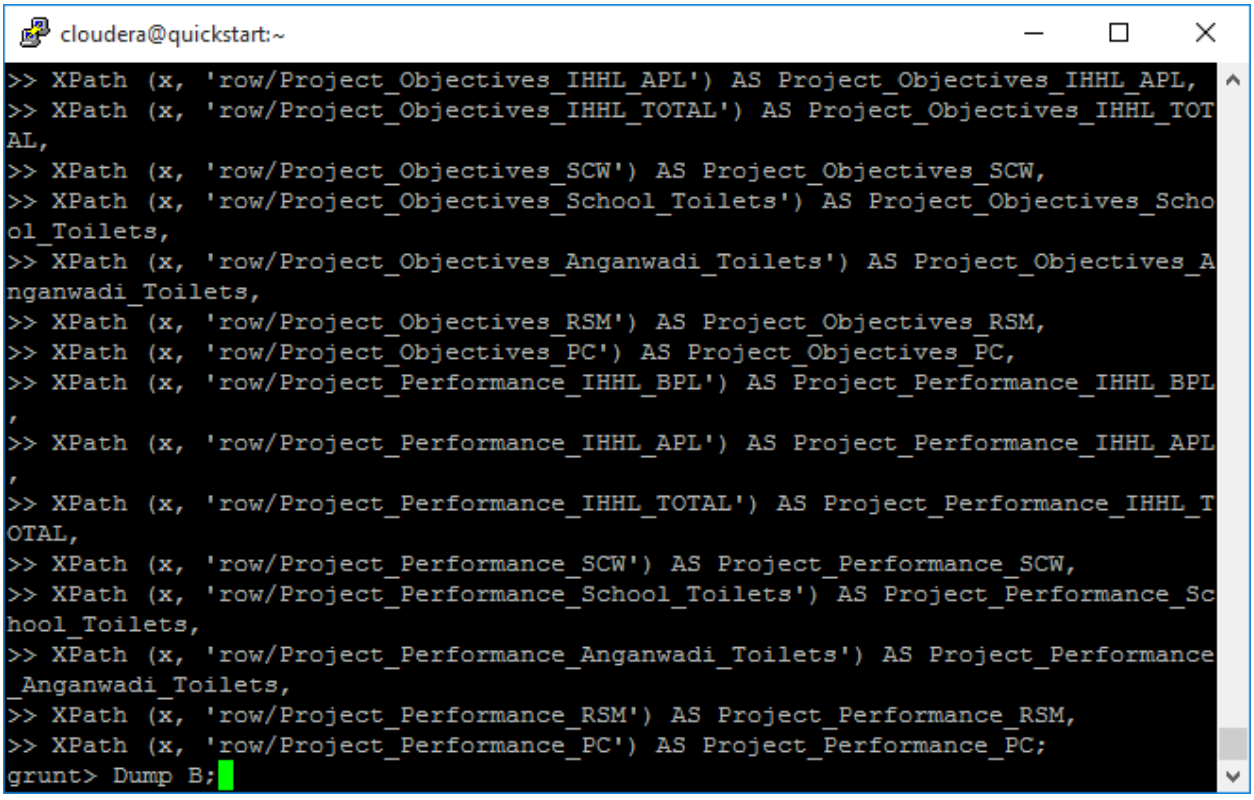




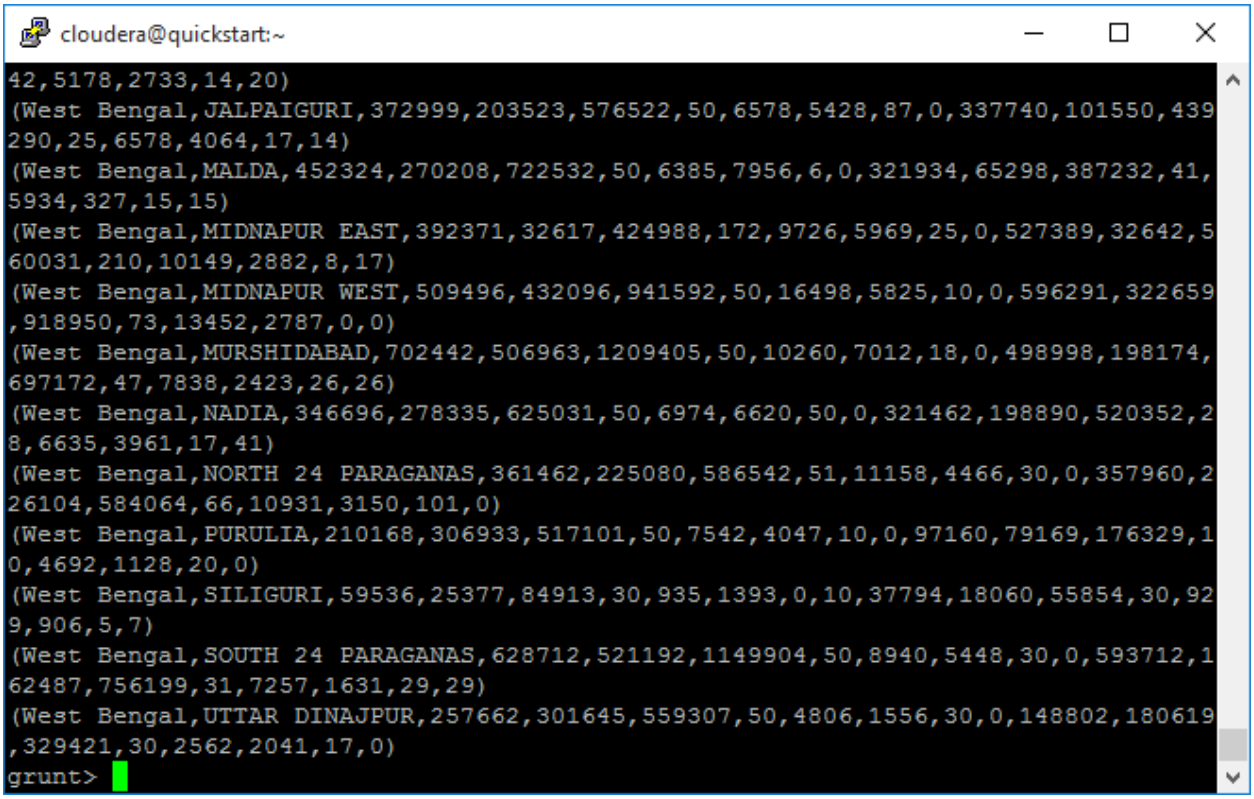
C. Screen shot for the output of the Problem Statement2:



After dump A, below is the result:



After dump B, below is the result:

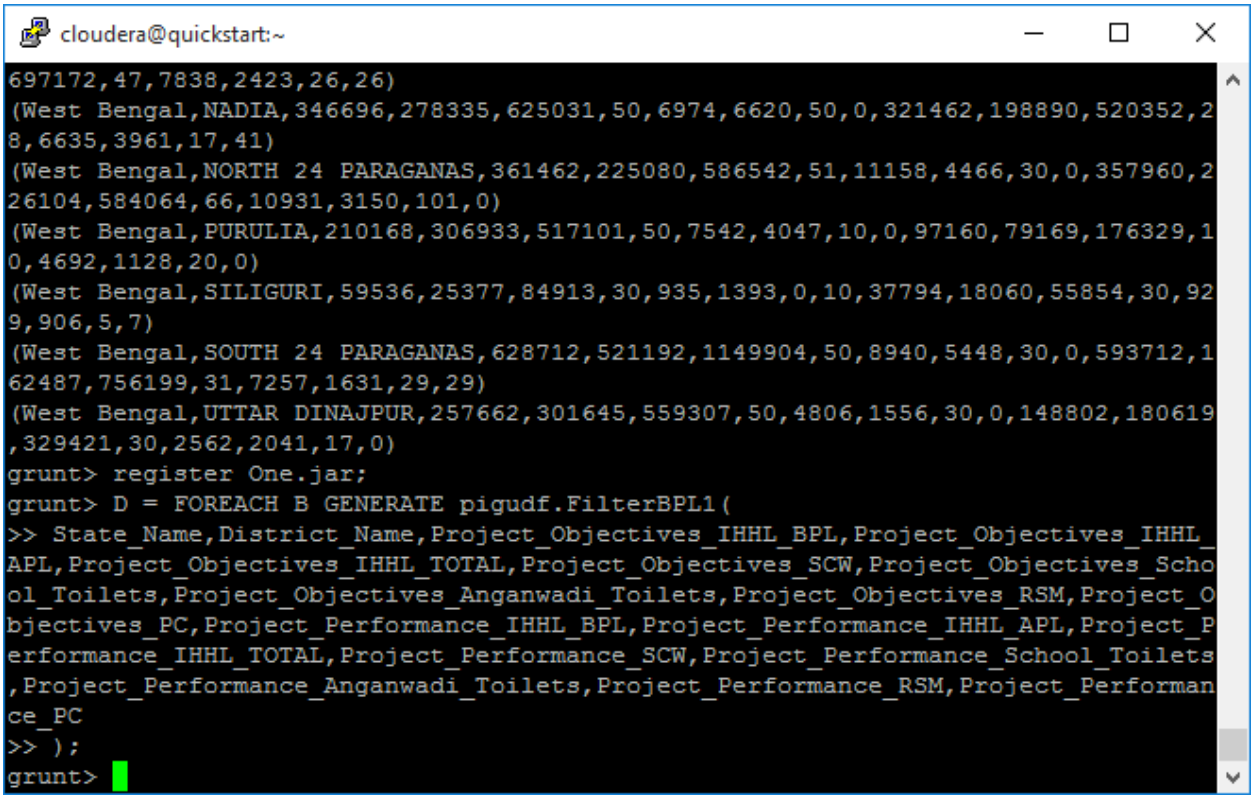


After Executing, register One.jar;

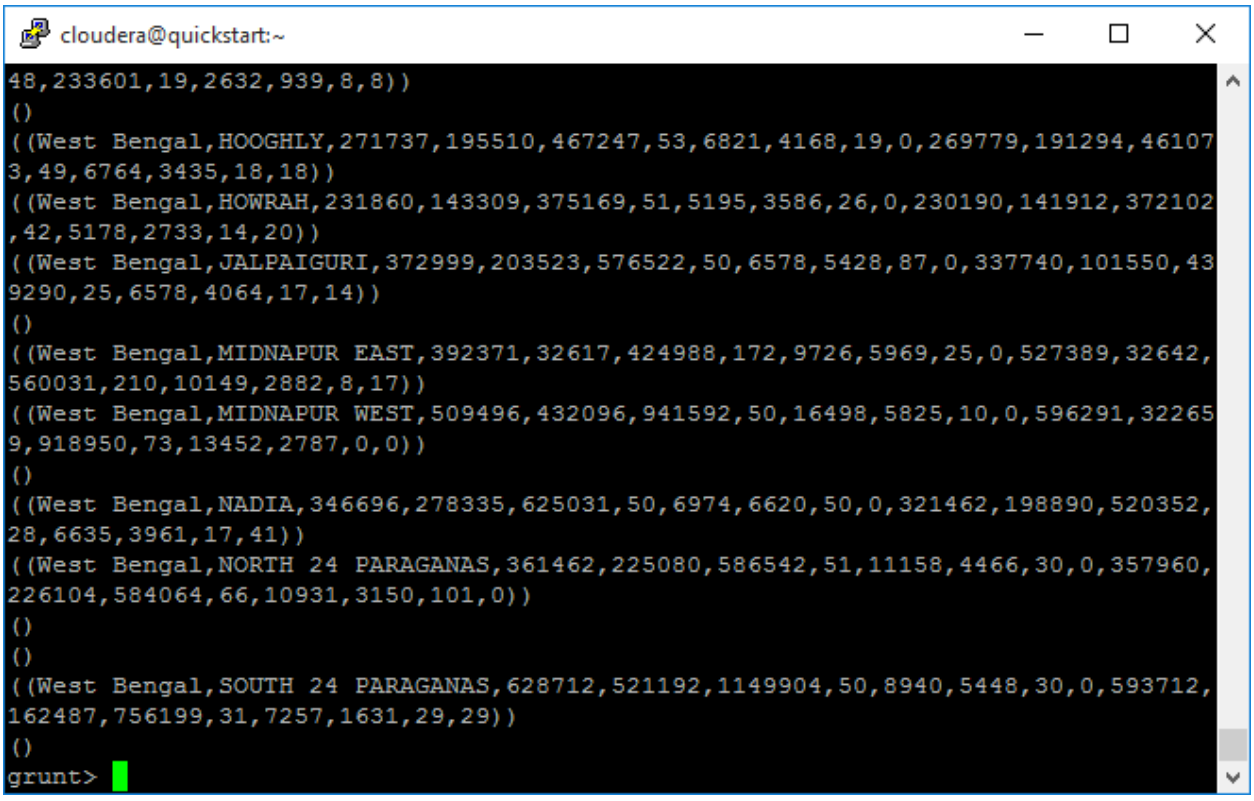
D = FOREACH B GENERATE pigudf.FilterBPL1(

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

);



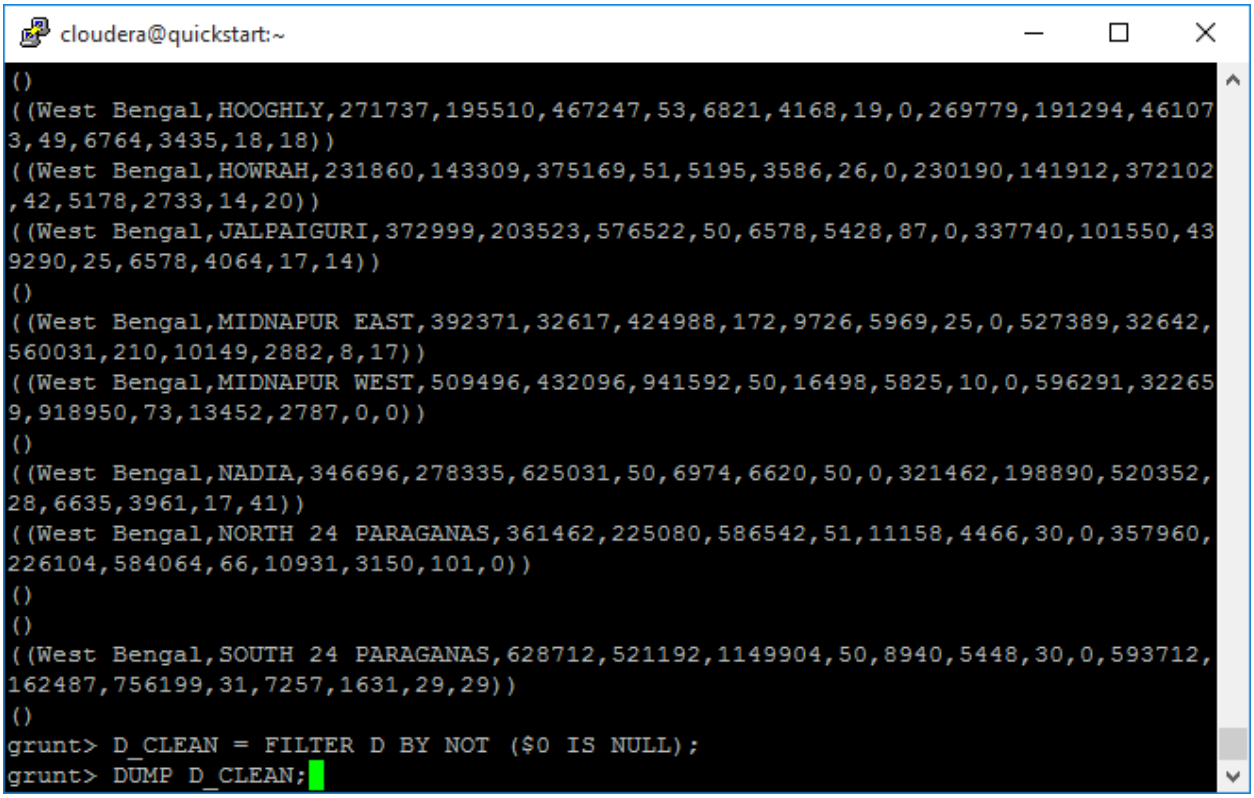
After dump D, below is the result:

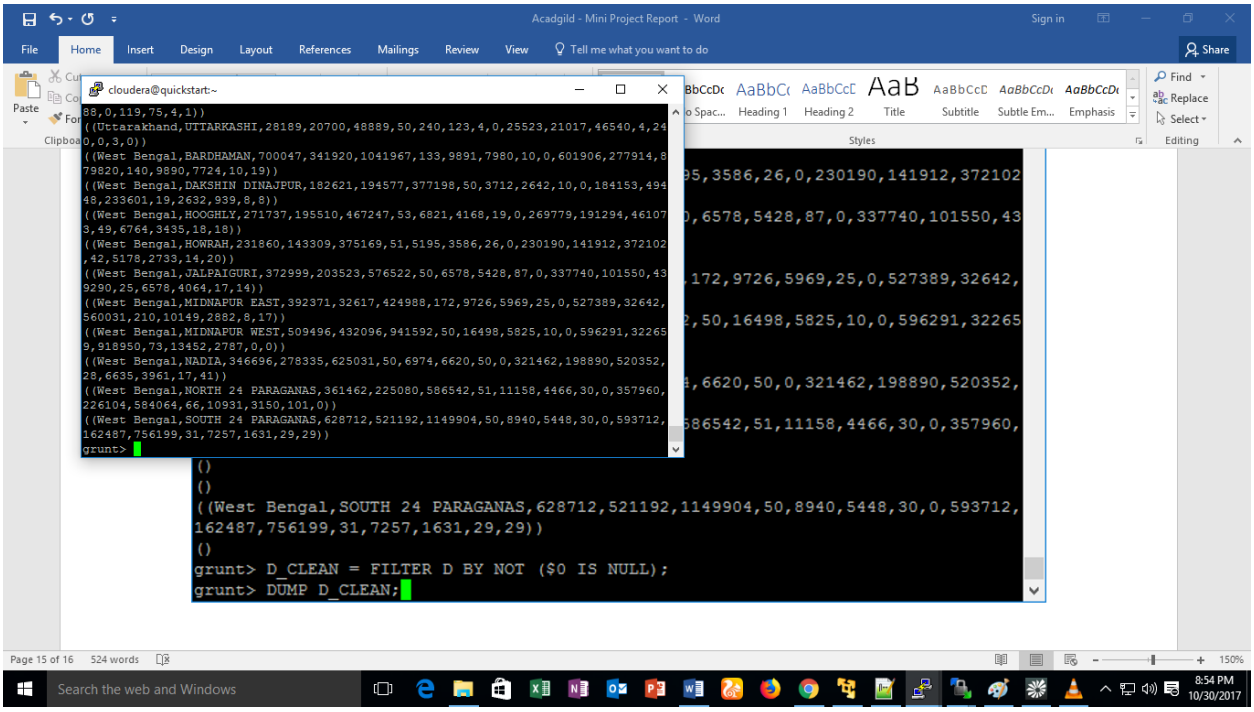


After executing the below command, below is the screen:

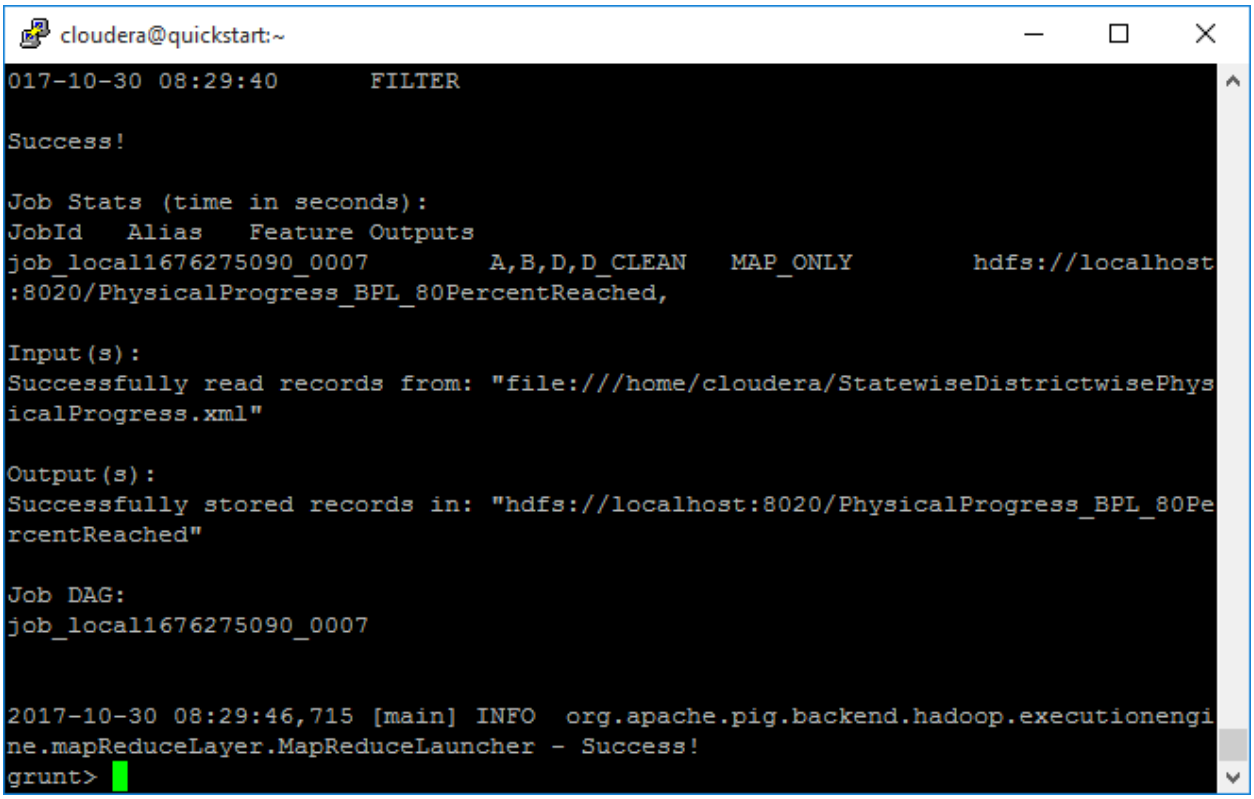
D_CLEAN = FILTER D BY NOT ($0 IS NULL);

DUMP D_CLEAN;

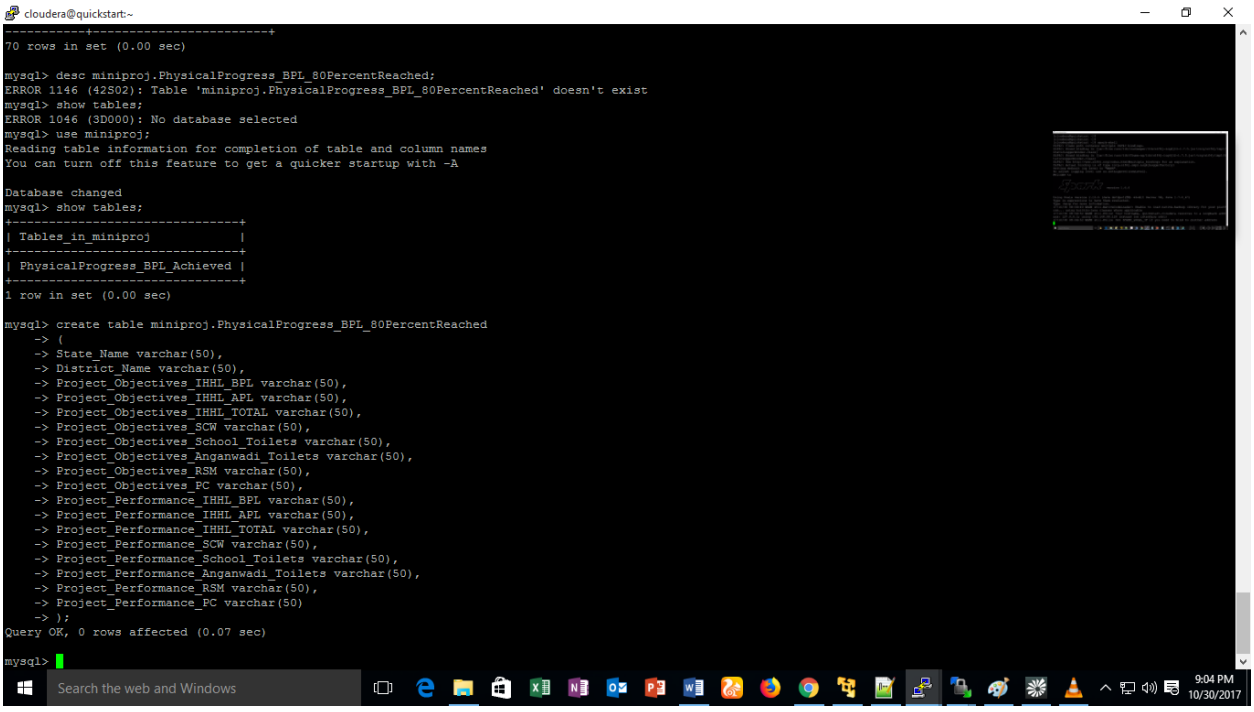




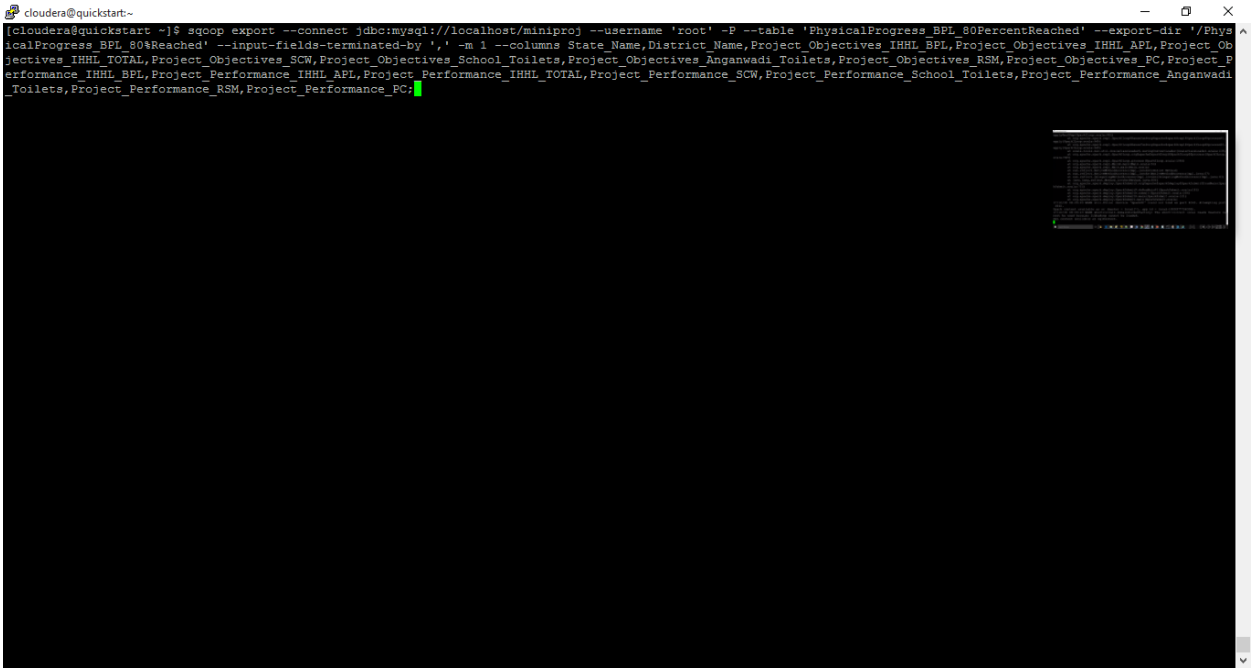
After executing- STORE D_CLEAN INTO 'hdfs://localhost:8020/PhysicalProgress_BPL_80%Reached/' USING PigStorage (',');



After following command in MYSQL, below is the result:

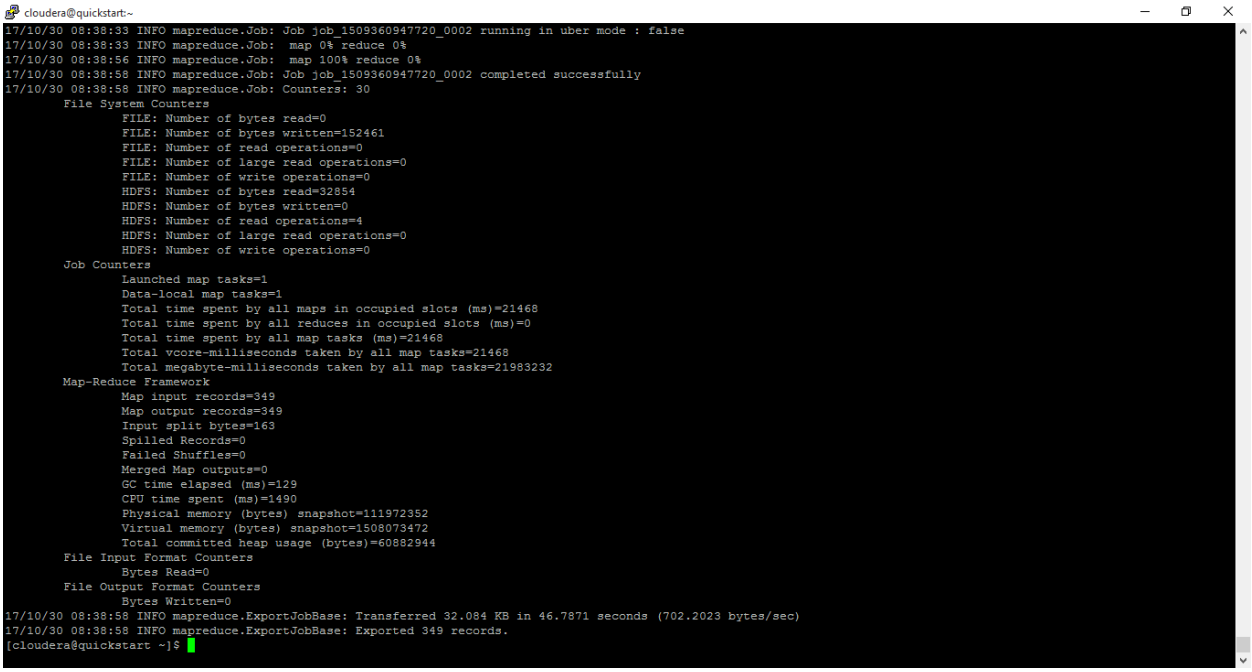


In following screen, below command is executed:

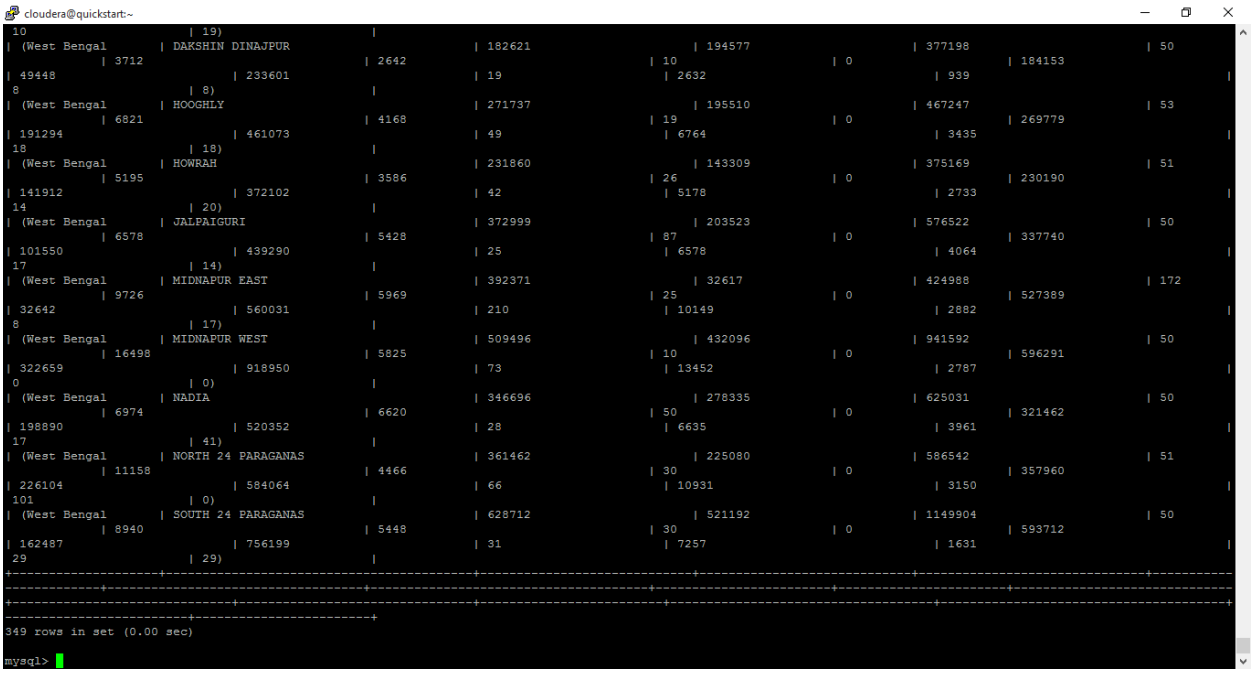


After executing the Linux command, show the below screen shot:

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;



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



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.